```
             YGFDAGYL E GPAD+VIF   +ER+I   FASK+SNSPFIG+KLKGV+ YTI +GE+VY
Sbjct: 361 YGFDAGYLAENGPADLVIFADKQERLITENFASKASNSPFIGNKLKGVVKYTIADGEVVY 420
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 662

A DNA sequence (GBSx0702) was identified in *S.agalactiae* <SEQ ID 2039> which encodes the amino acid sequence <SEQ ID 2040>. This protein is predicted to be orotate phosphoribosyltransferase PyrE (pyrE). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2214(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC95453 GB:AF068902 orotate phosphoribosyltransferase PyrE
            [Streptococcus pneumoniae]
 Identities = 152/208 (73%), Positives = 180/208 (86%)

Query: 1   MDLARQIAMELLDIQAVYLRPQQPFTWASGVKSPIYTDNRVTLSYPETRTLIENGFVKQI 60
            M LA+ IA  LL IQAVYL+P++PFTWASG+KSPIYTDNRVTL+YPETRTLIENGFV  I
Sbjct: 1   MTLAKDIASHLLKIQAVYLKPEEPFTWASGIKSPIYTDNRVTLAYPETRTLIENGFVDAI 60

Query: 61  QKHFPNVDIIAGTATAGIPHGAIIADKMNLPFAYIRSKAKDHGVGNQIEGRVYSGQKMVI 120
            ++ FP V+++IAGTATAGIPHGAIIADKMNLPFAYIRSK KDHG GNQIEGRV  GQKMV+
Sbjct: 61  KEAFPEVEVIAGTATAGIPHGAIIADKMNLPFAYIRSKPKDHGAGNQIEGRVAQGQKMVV 120

Query: 121 IEDLISTGGSVLEAVTAAQSQGIEVLGVVAIFTYQLAKAEQAFREADIPLVTLTDYNQLI 180
            +EDLISTGGSVLEAV AA+ +G +VLGVVAIF+YQL KA++ F +A + LVTL++Y++LI
Sbjct: 121 VEDLISTGGSVLEAVAAAKREGADVLGVVAIFSYQLPKADKNFADAGVKLVTLSNYSELI 180

Query: 181 KVAKVNGYITADQLVLLKKFKEDQMNWQ 208
             +A+  GYIT + L LLK+FKEDQ NWQ
Sbjct: 181 HLAQEEGYITPEGLDLLKRFKEDQENWQ 208
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2041> which encodes the amino acid sequence <SEQ ID 2042>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1612(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 158/208 (75%), Positives = 179/208 (85%)

Query: 1   MDLARQIAMELLDIQAVYLRPQQPFTWASGVKSPIYTDNRVTLSYPETRTLIENGFVKQI 60
            M LA QIA +LLDI+AVYL+P+ PFTWASG+KSPIYTDNRVTLSYP+TR LIENGFV+ I
Sbjct: 1   MTLASQIATQLLDIKAVYLKPEDPFTWASGIKSPIYTDNRVTLSYPKTRDLIENGFVETI 60

Query: 61  QKHFPNVDIIAGTATAGIPHGAIIADKMNLPFAYIRSKAKDHGVGNQIEGRVYSGQKMVI 120
            + HFP V+++IAGTATAGIPHGAIIADKM LPFAYIRSK KDHG GNQIEGRV  GQKMVI
Sbjct: 61  KAHFPEVEVIAGTATAGIPHGAIIADKMTLPFAYIRSKPKDHGAGNQIEGRVLKGQKMVI 120
```

```
Query: 121 IEDLISTGGSVLEAVTAAQSQGIEVLGVVAIFTYQLAKAEQAFREADIPLVTLTDYNQLI 180
           IEDLISTGGSVL+A  AA  +G +VLGVVAIFTY+L KA Q F+EA I L+TL++Y +LI
Sbjct: 121 IEDLISTGGSVLDAAAAASREGADVLGVVAIFTYELPKASQNFKEAGIKLITLSNYTELI 180

Query: 181 KVAKVNGYITADQLVLLKKFKEDQMNWQ 208
            VAK+ GYIT D L LLKKFKEDQ+NWQ
Sbjct: 181 AVAKLQGYITNDGLHLLKKFKEDQVNWQ 208
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 663

A DNA sequence (GBSx0703) was identified in *S.agalactiae* <SEQ ID 2043> which encodes the amino acid sequence <SEQ ID 2044>. This protein is predicted to be orotidine 5'-phosphate decarboxylase (pyrF). Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9829> which encodes amino acid sequence <SEQ ID 9830> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC95452 GB:AF068902 orotidine-5'-decarboxylase PyrF
           [Streptococcus pneumoniae]
 Identities = 149/231 (64%), Positives = 176/231 (75%), Gaps = 1/231 (0%)

Query: 19  MLEKCPIIALDFSDLASVTTFLEHFPKEELLFVKIGMELYYSEGPSIIRYIKSLGHRIFL 78
           M E  PIIALDF    +V  FL  FP EE L++K+GMELYY+ GP I+ Y+K LGH +FL
Sbjct: 1   MREHRPIIALDFPSFEAVKEFLALFPAEESLYLKVGMELYYAAGPEIVSYLKGLGHSVFL 60

Query: 79  DLKLHDIPNTVRSSMSVLAKLGIDMTNVHAAGGVEMMKAAREGLGKGPILLAVTQLTSTS 138
           DLKLHDIPNTV+S+M VL++LG+DMTNVHAAGGVEMMKAAREGLG    L+AVTQLTSTS
Sbjct: 61  DLKLHDIPNTVKSAMKVLSQLGVDMTNVHAAGGVEMMKAAREGLGSQAKLIAVTQLTSTS 120

Query: 139 QEQMQVDQHINLSVVDSVCHYAQKAQEAGLDGVVASAQEGMQIKKQTNEHFICLTPGIRP 198
           + QMQ  Q+I  S+ +SV HYA+K  EAGLDGVV SAQE   IK+ TN  FICLTPGIRP
Sbjct: 121 EAQMQEFQNIQTSLQESVIHYAKKTAEAGLDGVVCSAQEVQVIKQATNPDFICLTPGIRP 180

Query: 199 PQTNQLDDQKRTMTPEQARIVGADYIVVGRPITKAENPYQAYLEIKEEWNR 249
                + DQKR MTP  A  +G+DYIVVGRPIT+AE+P  AY  IK+EW +
Sbjct: 181 AGV-AVGDQKRVMTPADAYQIGSDYIVVGRPITQAEDPVAAYHAIKDEWTQ 230
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2045> which encodes the amino acid sequence <SEQ ID 2046>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1934(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 149/229 (65%), Positives = 180/229 (78%), Gaps = 1/229 (0%)

Query: 19  MLEKCPIIALDFSDLASVTTFLEHFPKEELLFVKIGMELYYSEGPSIIRYIKSLGHRIFL 78
            M E+ PIIALDFS       FL+ FP EE L+VKIGMELYY++GP I+RYIKSLGH +FL
Sbjct: 1   MKEERPIIALDFSSFEETKAFLDLFPAEEKLYVKIGMELYYAQGPDIVRYIKSLGHNVFL 60

Query: 79  DLKLHDIPNTVRSSMSVLAKLGIDMTNVHAAGGVEMMKAAREGLGKGPILLAVTQLTSTS 138
            DLKLHDIPNTVR++M+VL +L IDM  VHAAGGVEM+KAAREGLG+GP L+AVTQLTSTS
Sbjct: 61  DLKLHDIPNTVRAAMAVLKELDIDMATVHAAGGVEMLKAAREGLGQGPTLIAVTQLTSTS 120

Query: 139 QEQMQVDQHINLSVVDSVCHYAQKAQEAGLDGVVASAQEGMQIKKQTNEHFICLTPGIRP 198
            ++QM+ DQ+I  S+++SV HY++ A +A LDG V SAQE   IK  T   F CLTPGIRP
Sbjct: 121 EDQMRGDQNIQTSLLESVLHYSKGAAKAQLDGAVCSAQEVEAIKAVTPTGFTCLTPGIRP 180

Query: 199 PQTNQLDDQKRTMTPEQARIVGADYIVVGRPITKAENPYQAYLEIKEEW 247
              +N + DQKR MTP QAR +G+DYIVVGRPIT+A++P  AY  IK EW
Sbjct: 181 KGSN-IGDQKRVMTPNQARRIGSDYIVVGRPITQAKDPVAAYQAIKAEW 228
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 664

A DNA sequence (GBSx0704) was identified in *S.agalactiae* <SEQ ID 2047> which encodes the amino acid sequence <SEQ ID 2048> in others. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.70   Transmembrane  192 - 208 ( 190 - 211)
   INTEGRAL    Likelihood = -7.64   Transmembrane  226 - 242 ( 218 - 250)
   INTEGRAL    Likelihood = -3.77   Transmembrane  388 - 404 ( 378 - 404)
   INTEGRAL    Likelihood = -3.08   Transmembrane  293 - 309 ( 292 - 311)
   INTEGRAL    Likelihood = -2.87   Transmembrane  165 - 181 ( 162 - 182)
   INTEGRAL    Likelihood = -2.13   Transmembrane  267 - 283 ( 267 - 284)
   INTEGRAL    Likelihood = -0.90   Transmembrane  114 - 130 ( 114 - 130)
   INTEGRAL    Likelihood = -0.75   Transmembrane  318 - 334 ( 318 - 334)
   INTEGRAL    Likelihood = -0.53   Transmembrane  140 - 156 ( 140 - 156)

----- Final Results -----
               bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB03800 GB:AP001507 unknown conserved protein in others
         [Bacillus halodurans]
 Identities = 63/243 (25%), Positives = 120/243 (48%)

Query: 5   MSVVLRAGKLLIESGAEVYRVEDTMKHFAKALQIENFEAYVVSSSIIASGINRYGKQEAK 64
            M + + AG++++ +GAE YRVE+T++  AKA Q  N  ++V ++ I  S          +
Sbjct: 8   MDICMLAGEIMLINGAETYRVEETLERMAKAGQFRNVHSFVTTTGIFLSFEEEGAGDVMQ 67

Query: 65  VCNTDGVTANLGRLEAVNNLSRQIAKQDLVSPEEIVKQLDLIEHQKDYSLLVTLISYFCG 124
            +   D    +L ++  VN +SR+    ++ + E + K  ++ +   +YS L+   +
Sbjct: 68  MIRVDDRMQDLNKVTLVNQVSREFVNGEIDAAEALTKLQNIAKQPMNYSPLLLHTASGVA 127

Query: 125 AGSFSLALGSSLLDSFSAAVTGLILGYFLNLMESRIHTGFLLTILGSSVVALSANLLYFS 184
             G+FS   G +L D+  A + G +    +  ++S +   F    + +      A LL
Sbjct: 128 GGAFSYLFGGNLFDTLPAFIAGFVASMAVVHLQSYLKVRFFAEFMAAFTGGAVAILLVLI 187

Query: 185 GLGEHRSIIILGALMVMVPGAAFVNSVREFSQNNFSTGLALIMSALLICISISAGVAITI 244
            GLGE+   +I+G LM +VPG    N+VR+    +   G+       +  +SI+ G+A+ I
Sbjct: 188 GLGENVDQVIIGTLMPLVPGIPLTNAVRDLISGDLLAGVTRGAECFVTSLSIATGIALAI 247

Query: 245 EII 247
```

```
            ++
Sbjct: 248 ALL 250
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 665

A DNA sequence (GBSx0705) was identified in *S.agalactiae* <SEQ ID 2049> which encodes the amino acid sequence <SEQ ID 2050>. This protein is predicted to be ABC transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5134(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9353> which encodes amino acid sequence <SEQ ID 9354> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12571 GB:Z99108 similar to ABC transporter (ATP-binding
           protein) [Bacillus subtilis]
 Identities = 193/288 (67%), Positives = 231/288 (80%)

Query: 1   MNDVINIVYHVENQDLVRYSGDYTNFESVYAMKKAQLEAAYERQQKEIADLQDFVNRNKA 60
           +N VIN++YHVENQ+L RY GDY  F  VY +KK QLEAAY++QQ+E+A+L+DFV RNKA
Sbjct: 222 LNSVINLIYHVENQELTRYVGDYHQFMEVYEVKKQQLEAAYKKQQQEVAELKDFVARNKA 281

Query: 61  RVATRNMAMSRQKKLDKMDIIELQAEKPKPSFEFKESRTPGRFIFQAKDLQIGYDRALTK 120
           RV+TRNMAMSRQKKLDKMD+IEL AEKPKP F FK +RT G+ IF+ KDL IGYD  L++
Sbjct: 282 RVSTRNMAMSRQKKLDKMDMIELAAEKPKPEFHFKPARTSGKLIFETKDLVIGYDSPLSR 341

Query: 121 PLNLTFERNQKIAIVGANGIGKTTLLKSLLGIIPPISGNVERGDFIDLGYFEQEVPGGNR 180
           PLNL  ER QKIA+ GANGIGKTTLLKSLLG I P+ G+VERG+ I  GYFEQEV   N
Sbjct: 342 PLNLRMERGQKIALYGANGIGKTTLLKSLLGEIQPLEGSVERGEHIYTGYFEQEVKETNN 401

Query: 181 QTPLEAVWDAFPALNQAEVRAALARCGLTSKHIESQIQVLSGGEQSKVRFCLLMNRENNV 240
            T +E VW  FP+  Q E+RAA A+CGLT+KHIES++ VLSGGE++KVR C L+N E N+
Sbjct: 402 NTCIEEVWSEFPSYTQYEIRAAPAKCGLTTKHIESRVSVLSGGEKAKVRLCKLINSETNL 461

Query: 241 LVLDEPTNHLDVDAKDELKRALKAYKGSILMVCHEPDFYEGWMDDVWD 288
           LVLDEPTNHLD DAK+ELKRALK YKGSIL++ HEPDFY     + W+
Sbjct: 462 LVLDEPTNHLDADAKEELKRALKEYKGSILLISHEPDFYMDIATETWN 509
 Identities = 56/219 (25%), Positives = 97/219 (43%), Gaps = 44/219 (20%)

Query: 104 IFQAKDLQIGY-DRALTKPLNLTFERNQKIAIVGANGIGKTTLLKSLLGIIPPISGNVER 162
           I   KDL  G+ DRA+   ++    + + + ++GANG GK+T +  + G + P  G VE
Sbjct: 3   ILSVKDLSHGFGDRAIFNNVSFRLLKGEHVGLIGANGEGKSTFMNIITGKLEPDEGKVEW 62

Query: 163 GDFIDLGYFEQEVPGGNRQTPLEAVWDAFPALNQAE------------------------ 198
              + +GY +Q       ++  + + DAF  L   E
Sbjct: 63  SKNVRVGYLDQHTVLEKGKSIRDVLKDAFHYLFAMEEEMNEIYNKMGEADPDELEKLLEE 122

Query: 199 ---VRAALAR----------------CGLTSKHIESQIQVLSGGEQSKVRFCLLMNRENN 239
              ++ AL                   GL+   +E  +  LSGG+++KV    L+  +
Sbjct: 123 VGVIQDALTNNDFYVIDSKVEEIARGLGLSDIGLERDVTDLSGGQRTKVLLAKLLLEKPE 182
```

```
Query: 240 VLVLDEPTNHLDVDAKDELKRALKAYKGSILMVCHEPDF 278
           +L+LDEPTN+LD    + LKR L+ Y+ + +++ H+  F
Sbjct: 183 ILLLDEPTNYLDEQHIEWLKRYLQEYENAFILISHDIPF 221
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2051> which encodes the amino acid sequence <SEQ ID 2052>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2794(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 246/294 (83%), Positives = 274/294 (92%), Gaps = 1/294 (0%)

Query: 1   MNDVINIVYHVENQDLVRYSGDYTNFESVYAMKKAQLEAAYERQQKEIADLQDFVNRNKA 60
           +NDVINIVYHVENQ LVRY+GDY  F++VY MK++QLEAAYERQQKEIA+LQDFVNRNKA
Sbjct: 233 LNDVINIVYHVENQSLVRYTGDYYQFQAVYEMKQSQLEAAYERQQKEIANLQDFVNRNKA 292

Query: 61  RVATRNMAMSRQKKLDKMDIIELQAEKPKPSFEFKESRTPGRFIFQAKDLQIGYDRALTK 120
           RVATRNMAMSRQKKLDKMDIIELQAEKPKP+FEFK++RTP RFIFQ K+L IGYD  LTK
Sbjct: 293 RVATRNMAMSRQKKLDKMDIIELQAEKPKPNFEFKQARTPSRFIFQTKNLVIGYDYPLTK 352

Query: 121 -PLNLTFERNQKIAIVGANGIGKTTLLKSLLGIIPPISGNVERGDFIDLGYFEQEVPGGN 179
            PLN+TFERNQKIAIVGANGIGK+TLLKSLLG+I P+ G++  GDF+++GYFEQEV G N
Sbjct: 353 EPLNITFERNQKIAIVGANGIGKSTLLKSLLGVIEPLEGHIVTGDFLEVGYFEQEVTGVN 412

Query: 180 RQTPLEAVWDAFPALNQAEVRAALARCGLTSKHIESQIQVLSGGEQSKVRFCLLMNRENN 239
           RQTPLE VWDAFPALNQAEVRAALARCGLTSKHIESQIQVLSGGEQ+KVRFCLLMNRENN
Sbjct: 413 RQTPLEVVWDAFPALNQAEVRAALARCGLTSKHIESQIQVLSGGEQAKVRFCLLMNRENN 472

Query: 240 VLVLDEPTNHLDVDAKDELKRALKAYKGSILMVCHEPDFYEGWMDDVWDFNQLS 293
           VL+LDEPTNHLD+DAK+ELKRALKAYKGSILMVCHEPDFY GW+ D WDF++L+
Sbjct: 473 VLILDEPTNHLDIDAKNELKRALKAYKGSILMVCHEPDFYNGWVTDTWDFSKLT 526
 Identities = 60/218 (27%), Positives = 102/218 (46%), Gaps = 43/218 (19%)

Query: 104 IFQAKDLQIGY-DRALTKPLNLTFERNQKIAIVGANGIGKTTLLKSLLGIIPPISGNVER 162
           I + K L  G+ DRA+ + ++    + + I +VGANG GK+T +  + G + P  G VE
Sbjct: 15  ILEVKQLSHGFGDRAIFENVSFRLLKGEHIGLVGANGEGKSTFMSIVTGHLQPDEGKVEW 74

Query: 163 GDFIDLGYFEQEVPGGNRQTPLEAVWDAFPALNQAEVR-----AALA------------- 204
             ++  GY +Q     + QT  + +  AF  L + E R     A++A
Sbjct: 75  SKYVTAGYLDQHTVLESGQTVRDVLRTAFDELFKTENRINEIYASMADDKADIAVLMEEV 134

Query: 205 ------------------------RCGLTSKHIESQIQVLSGGEQSKVRFCLLMNRENNV 240
                                     G+    +ES +  LSGG+++KV    L+  + ++
Sbjct: 135 GELQDRLESRDFYTLDAKIDEVARALGVMDFGMESDVTSLSGGQRTKVLLAKLLLEKPDI 194

Query: 241 LVLDEPTNHLDVDAKDELKRALKAYKGSILMVCHEPDF 278
           L+LDEPTNHLD +  + LKR L+ Y+ + +++ H+  F
Sbjct: 195 LLLDEPTNHLDAEHIEWLKRYLQHYENAFVLISHDISF 232
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 666

A DNA sequence (GBSx0706) was identified in *S.agalactiae* <SEQ ID 2053> which encodes the amino acid sequence <SEQ ID 2054>. This protein is predicted to be lipoprotein Nlp1 precursor (pstS). Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2637(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14429 GB:Z99116 alternate gene name: yzmB~similar to
           phosphate ABC transporter (binding protein) [Bacillus subtilis]
 Identities = 42/62 (67%), Positives = 49/62 (78%)

Query: 15  SITSVGSTALQPLVEAAADEFGKTNLGKTINVQGGGSGTGLSQVQSGAVQIGNSDLFAEE 74
            S+T  GS+A+QPLV AAA++F + N    I VQ GGSGTGLSQV  GAVQIGNSD+FAEE
Sbjct: 45  SLTISGSSAMQPLVLAAAEKFMEENPDADIQVQAGGSGTGLSQVSEGAVQIGNSDVFAEE 104

Query: 75  KE 76
            KE
Sbjct: 105 KE 106
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1695> which encodes the amino acid sequence <SEQ ID 1696>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 63/74 (85%), Positives = 71/74 (95%)

Query: 3  LSGCANWIDKGQSITSVGSTALQPLVEAAADEFGKTNLGKTINVQGGGSGTGLSQVQSGA 62
           LS C++WIDKG+SIT+VGSTALQPLVEA ADEFG +NLGKT+NVQGGGSGTGLSQVQSGA
Sbjct: 20 LSACSSWIDKGESITAVGSTALQPLVEAVADEFGSSNLGKTVNVQGGGSGTGLSQVQSGA 79

Query: 63 VQIGNSDLFAEEKE 76
           VQIGNSD+FAEEK+
Sbjct: 80 VQIGNSDVFAEEKD 93
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 667

A DNA sequence (GBSx0707) was identified in *S.agalactiae* <SEQ ID 2055> which encodes the amino acid sequence <SEQ ID 2056>. This protein is predicted to be lipoprotein Nlp1 precursor (pstS). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have an uncleavable N-term signal seq
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9343> which encodes amino acid sequence <SEQ ID 9344> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14429 GB:Z99116 alternate gene name: yzmB~similar to
           phosphate ABC transporter (binding protein) [Bacillus subtilis]
 Identities = 95/184 (51%), Positives = 126/184 (67%), Gaps = 1/184 (0%)

Query: 3   DHQVAVAGLAVIVNKKVNVKNLTTHQLRDIFAGKIKNWKEVGGQDLDISIINRAASSGSR 62
           DHQVAV G+A  VN    VK+++  +L+ IF GKIKNWKE+GG+D  I+++NR  SSG+R
Sbjct: 115 DHQVAVVGMAAAVNPDAGVKDISKDELKKIFTGKIKNWKELGGKDQKITLVNRPDSSGTR 174

Query: 63  ATFDNTIMGNVAPIQSQEQDSNGMVKSIVSQTPGAISYLAFAYV-DKSVGTLKLNGFAPT 121
           ATF    +    P +   +DS+  VK I++ TPGAI YLAF+Y+ D  V  L ++G  P
Sbjct: 175 ATFVKYALDGAEPAEGITEDSSNTVKKIIADTPGAIGYLAFSYLTDDKVTALSIDGVKPE 234

Query: 122 AKNVTTDNWKLWSYEHMYTKGNETGLTKEFLDYMKSDKVQSSIVQHMGYISINDMKVVKD 181
           AKNV T  + +W+Y+H YTKG  TGL KEFLDY+KS+ +Q SIV   GYI + DMKV +D
Sbjct: 235 AKNVATGEYPIWAYQHSYTKGEATGLAKEFLDYLKSEDIQKSIVTDQGYIPVTDMKVTRD 294

Query: 182 AEGK 185
           A GK
Sbjct: 295 ANGK 298
```

There is also homology to SEQ ID 1696.

SEQ ID 9344 (GBS659) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 135 (lane 2 & 3; MW 60kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 135 (lane 5-7; MW 35kDa) and in Figure 178 (lane 11; MW 35kDa).

GBS659-His was purified as shown in Figure 228, lane 6-8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 668

A DNA sequence (GBSx0708) was identified in *S.agalactiae* <SEQ ID 2057> which encodes the amino acid sequence <SEQ ID 2058>. This protein is predicted to be phosphate transporter permease PstC (pstC-2). Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-15.50   Transmembrane   35 -  51 (  27 -  61)
   INTEGRAL    Likelihood = -7.64   Transmembrane  167 - 183 ( 154 - 186)
   INTEGRAL    Likelihood = -6.37   Transmembrane  282 - 298 ( 277 - 302)
   INTEGRAL    Likelihood = -5.52   Transmembrane   85 - 101 (  81 - 116)
   INTEGRAL    Likelihood = -3.24   Transmembrane  133 - 149 ( 131 - 155)

----- Final Results -----
               bacterial membrane --- Certainty=0.7198(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8635> which encodes amino acid sequence <SEQ ID 8636> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
SRCFLG: 0
McG: Length  of  UR:    5
     Peak Value of UR:  -0.12
     Net Charge of CR: 2
McG: Discrim Score:    -16.22
GvH: Signal Score (-7.5): -4.26
     Possible site: 41
>>> Seems to have no N-terminal signal sequence
Amino Acid Composition: calculated from 1
ALOM program   count: 5 value: -15.50 threshold:  0.0
   INTEGRAL    Likelihood =-15.50   Transmembrane   29 -  45 (  21 -  55)
   INTEGRAL    Likelihood = -7.64   Transmembrane  161 - 177 ( 148 - 180)
   INTEGRAL    Likelihood = -6.37   Transmembrane  276 - 292 ( 271 - 296)
   INTEGRAL    Likelihood = -5.52   Transmembrane   79 -  95 (  75 - 110)
   INTEGRAL    Likelihood = -3.24   Transmembrane  127 - 143 ( 125 - 149)
   PERIPHERAL  Likelihood =  0.69     205
 modified ALOM score:   3.60
icml HYPID: 7  CFP: 0.720

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.7198(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14428 GB:Z99116 alternate gene name: yzmC~similar to
           phosphate ABC transporter (permease) [Bacillus subtilis]
 Identities = 145/303 (47%), Positives = 209/303 (68%), Gaps = 4/303 (1%)

Query: 8   KNQELAKKLTSPSKNSRLEKFGKGITFLSLALIVFIVAM-ILIFVAQKGLSTFFVDGVKL 66
           +N  ++++L S  +N +L++    +   + ALI+   ++ I IF+  KGL +F V+GV
Sbjct: 6   ENMSVSERLISSRQNRQLDEVRGRMIVTACALIMIAASVAITIFLGVKGLQSFLVNGVSP 65

Query: 67  TDFLFNTKWEP--SAKSFGAFPMIAGSFIVTILSAIIATPFAIGAAVFMTEISPKYGSKI 124
            +FL +  W P  S   +G  P I GSF VTILSA+IA P  I   +FMTEI+P +G K+
Sbjct: 66  IEFLTSLNWNPTDSDPKYGVLPFIFGSFAVTILSALIAAPLGIAGPIFMTEIAPNWGKKV 125

Query: 125 LQPAVELLVGIPSVVYGFIGLQIIVPFVRSI-FGGTGFGILSGVCVLFVMILPTVTFMTV 183
           LQP +ELLVGIPSVVYGFIGL ++VPF+      GTG  +L+G  VL VMILPT+T ++
Sbjct: 126 LQPVIELLVGIPSVVYGFIGLTVLVPFIAQFKSSGTGHSLLAGTIVLSVMILPTITSISA 185

Query: 184 DSLRAVPRHYKEASLAMGATRWQTIWRVILNAARPGIFTAIVFGMARAFGEALAIQMVVG 243
           D++ ++P+  +E S A+GATRWQTI +V++ AA P + TA+V GMARAFGEALA+QMV+G
Sbjct: 186 DAMASLPKSLREGSYALGATRWQTIRKVLVPAAFPTLMTAVVLGMARAFGEALAVQMVIG 245

Query: 244 NSAILPTSLTTPAATLTSVLTMGIGNTVMGTVQNNVLWSLALVLLIMSLAFNTVIKLITR 303
           N+ +LP S    A TLT+++T+ +G+T  G+V+NN LWS+ LVLL+MS  F  +I+ ++
Sbjct: 246 NTRVLPESPFDTAGTLTTIITLNMGHTTYGSVENNTLWSMGLVLLVMSFLFILLIRYLSS 305

Query: 304 EGK 306
             K
Sbjct: 306 RRK 308
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1691> which encodes the amino acid sequence <SEQ ID 1692>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-17.25   Transmembrane   29 -  45 (  21 -  55)
```

```
    INTEGRAL    Likelihood = -7.22   Transmembrane  162 - 178 ( 154 - 184)
    INTEGRAL    Likelihood = -5.57   Transmembrane  282 - 298 ( 277 - 302)
    INTEGRAL    Likelihood = -5.41   Transmembrane   96 - 112 (  81 - 116)
    INTEGRAL    Likelihood = -3.08   Transmembrane  133 - 149 ( 131 - 152)

----- Final Results -----
            bacterial membrane --- Certainty=0.7899(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 266/311 (85%), Positives = 290/311 (92%), Gaps = 6/311 (1%)

Query: 7   MKNQELAKKLTSPSKNSRLEKFGKGITFLSLALIVFIVAMILIFVAQKGLSTFFVDGVKL 66
           M+NQELAKKL SPSKNSRLE FG+ ITFL LALIVFIVAMILIFVAQKGLSTFFVD V L
Sbjct: 1   MENQELAKKLASPSKNSRLETFGRTITFLCLALIVFIVAMILIFVAQKGLSTFFVDKVNL 60

Query: 67  TDFLFNTKWEPSAKS------FGAFPMIAGSFIVTILSAIIATPFAIGAAVFMTEISPKY 120
            DFLF  +W+PS K+       GA PMI GSF+VTILSAIIATPFAIGAAVFMTEISPKY
Sbjct: 61  FDFLFGKEWQPSVKNAAGIPYLGALPMITGSFLVTILSAIIATPFAIGAAVFMTEISPKY 120

Query: 121 GSKILQPAVELLVGIPSVVYGFIGLQIIVPFVRSIFGGTGFGILSGVCVLFVMILPTVTF 180
           G+K+LQPAVELLVGIPSVVYGFIGLQ+IVPF+RSIFGGTGFGILSGVCVLFVMILPTVTF
Sbjct: 121 GAKLLQPAVELLVGIPSVVYGFIGLQVIVPFMRSIFGGTGFGILSGVCVLFVMILPTVTF 180

Query: 181 MTVDSLRAVPRHYKEASLAMGATRWQTIWRVILNAARPGIFTAIVFGMARAFGEALAIQM 240
           MT DSLRAVPRHY+EAS+AMGATRWQTIWRV+LNAARPGIFTA++FGMARAFGEALAIQM
Sbjct: 181 MTTDSLRAVPRHYREASMAMGATRWQTIWRVVLNAARPGIFTAVIFGMARAFGEALAIQM 240

Query: 241 VVGNSAILPTSLTTPAATLTSVLTMGIGNTVMGTVQNNVLWSLALVLLIMSLAFNTVIKL 300
           VVGNSA++P+SLTTPAATLTSVLTMGIGNTVMGTVQNNVLWSLALVLL+MSLAFN+++KL
Sbjct: 241 VVGNSAVMPSSLTTPAATLTSVLTMGIGNTVMGTVQNNVLWSLALVLLLMSLAFNSLVKL 300

Query: 301 ITREGKKNYER 311
           IT+E K+NYER
Sbjct: 301 ITKERKRNYER 311
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 669

A DNA sequence (GBSx0709) was identified in *S.agalactiae* <SEQ ID 2059> which encodes the amino acid sequence <SEQ ID 2060>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
           bacterial cytoplasm --- Certainty=0.2469(Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 670

A DNA sequence (GBSx0710) was identified in *S.agalactiae* <SEQ ID 2061> which encodes the amino acid sequence <SEQ ID 2062>. This protein is predicted to be probable abc transporter permease protein in soda-comga intergenic reg. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.24   Transmembrane   20 -  36 (  19 -  41)
   INTEGRAL    Likelihood = -8.28   Transmembrane   66 -  82 (  57 -  88)
   INTEGRAL    Likelihood = -6.90   Transmembrane  260 - 276 ( 258 - 285)
   INTEGRAL    Likelihood = -5.47   Transmembrane  109 - 125 ( 106 - 129)
   INTEGRAL    Likelihood = -2.87   Transmembrane  181 - 197 ( 178 - 198)

----- Final Results -----
               bacterial membrane --- Certainty=0.4694(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14427 GB:Z99116 alternate gene name: yzmD~similar to
           phosphate ABC transporter (permease) [Bacillus subtilis]
 Identities = 157/294 (53%), Positives = 225/294 (76%)

Query: 1   MNAKKADKLATTILYSIAAIIVTILASLLIFILVRGLPHVSWSFLTGKSSSYEAGGGIGI 60
           MN K  DKLAT +    AAII  IL  L  +I++ G+  +S+ F+T KSS+  AGGGI
Sbjct: 1   MNRKITDKLATGMFGLCAAIIAAILVGLFSYIIINGVSQLSFQFITTKSSAIAAGGGIRD 60

Query: 61  QLYNSFFLLIVTLIISIPLSLGAGIYLSEYAKKGRLTNFVRTCIEILSSLPSVVVGLFGY 120
           QL+NSF++L +T++I+IPL +G G++++EYA   ++T+F+RTCIE+LSSLPS+V+G+FG
Sbjct: 61  QLFNSFYILFITMLITIPLGVGGGVFMAEYAPNNKVTDFIRTCIEVLSSLPSIVIGMFGL 120

Query: 121 LIFVVQFQYGFSIISGALALTVFNLPQMTRSVEDSLQNVHHTQREAGLALGISRWETVIY 180
           L+FV    +G++II GALALTVFNLP M R  ED++++V    +EA LALG+SRW TV
Sbjct: 121 LMFVNLTGWGYTIIGGALALTVFNLPVMVRVTEDAIRSVPKDLKEASLALGVSRWHTVKT 180

Query: 181 VVVPEALPSIVTGVVLASGRIFGEAAALIYTAGQSAPALDWSNWNVLSVTSPISIFRQAE 240
           V++P A+PSI+TG +LASGR+FGEAAAL++TAG + P L+++ WN  S TSP++IFR AE
Sbjct: 181 VLIPSAIPSIITGAILASGRVFGEAAALLFTAGLTTPRLNFTEWNPFSETSPLNIFRPAE 240

Query: 241 TLAVHIWKVNSEGTIPDATQVSAGSAAVLLVVILIFNLSARSIGKKLHSKLTSS 294
           TLAVHIW VN++G IPDA  ++ G + VL++ +L+FNL+AR +G  ++ KLT++
Sbjct: 241 TLAVHIWNVNTQGMIPDAEAIANGGSPVLVISVLVFNLAARWLGTMIYKKLTAN 294
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1685> which encodes the amino acid sequence <SEQ ID 1686>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.89   Transmembrane   17 -  33 (   8 -  40)
   INTEGRAL    Likelihood =-10.19   Transmembrane  260 - 276 ( 257 - 285)
   INTEGRAL    Likelihood = -5.89   Transmembrane   66 -  82 (  57 -  87)
   INTEGRAL    Likelihood = -5.47   Transmembrane  109 - 125 ( 106 - 129)
   INTEGRAL    Likelihood = -2.02   Transmembrane  181 - 197 ( 180 - 197)

----- Final Results -----
               bacterial membrane --- Certainty=0.5755(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 257/294 (87%), Positives = 278/294 (94%)

Query: 1   MNAKKADKLATTILYSIAAIIVTILASLLIFILVRGLPHVSWSFLTGKSSSYEAGGGIGI 60
```

```
           MNAKK DK+AT  LY+IA IIV ILASL+++ILVRGLPH+SWSFLTGKSSSYEAGGGIGI
Sbjct: 1   MNAKKVDKVATGTLYTIAGIIVAILASLILYILVRGLPHISWSFLTGKSSSYEAGGGIGI 60

Query: 61  QLYNSFFLLIVTLIISIPLSLGAGIYLSEYAKKGRLTNFVRTCIEILSSLPSVVVGLFGY 120
           QLYNSFFLLIVTLIISIPLS GAGIYL+EYAKKG +TNF+RTCIEILSSLPSVVVGLFGY
Sbjct: 61  QLYNSFFLLIVTLIISIPLSTGAGIYLAEYAKKGPVTNFIRTCIEILSSLPSVVVGLFGY 120

Query: 121 LIFVVQFQYGFSIISGALALTVFNLPQMTRSVEDSLQNVHHTQREAGLALGISRWETVIY 180
           LIFVVQF+YGFSIISGALALTVFNLPQMTR+VEDSL +VHHTQREAGLALG+SRWETV Y
Sbjct: 121 LIFVVQFEYGFSIISGALALTVFNLPQMTRNVEDSLLHVHHTQREAGLALGLSRWETVFY 180

Query: 181 VVVPEALPSIVTGVVLASGRIFGEAAALIYTAGQSAPALDWSNWNVLSVTSPISIFRQAE 240
           VV+PEALP +VTG+VLASGRIFGEAAALIYTAGQSAPALDWSNWN LSVTSPISIFRQ+E
Sbjct: 181 VVIPEALPGMVTGIVLASGRIFGEAAALIYTAGQSAPALDWSNWNPLSVTSPISIFRQSE 240

Query: 241 TLAVHIWKVNSEGTIPDATQVSAGSAAVLLVVILIFNLSARSIGKKLHSKLTSS 294
           TLAVHIWKVNSEGTIPDAT VSAGSAAVLL+ ILIFN SA  IGKKLHSK+T++
Sbjct: 241 TLAVHIWKVNSEGTIPDATLVSAGSAAVLLIFILIFNFSAHFIGKKLHSKMTAA 294
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 671

A DNA sequence (GBSx0711) was identified in *S.agalactiae* <SEQ ID 2063> which encodes the amino acid sequence <SEQ ID 2064>. This protein is predicted to be phosphate ABC transporter, ATP-binding protein (pstB) (pstB-2). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4506(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB99016 GB:U67544 phosphate specific transport complex
           component (pstB) [Methanococcus jannaschii]
 Identities = 154/247 (62%), Positives = 204/247 (82%)

Query: 21  LTTKDLHVYYGEKEAIKGIDMQFEKNKITALIGPSGCGKSTYLRSLNRMNDTIDIARVTG 80
           + TK+L+++YGEK+A+  I++   +NKITALIGPSGCGKST+LR LNR+ND I   R+ G
Sbjct: 6   METKNLNLWYGEKQALFDINLPIYENKITALIGPSGCGKSTFLRCLNRLNDLIPNVRIEG 65

Query: 81  QIMYEGIDVNAQDINVYEMRKHIGMVFQRPNPFAKSIYKNITFAYERAGVKDKKFLDEVV 140
           +++ +G ++  +D++VYE+RK +GMVFQ+PNPFA SIY N+ F     G+KDKK LD++V
Sbjct: 66  EVLLDGKNIYDKDVDVYELRKRVGMVFQKPNPFAMSIYDNVAFGPRIHGIKDKKELDKIV 125

Query: 141 ETSLKQAALWDQVKDDLHKSAFTLSGGQQQRLCIARAIAVKPEILLMDEPASALDPIATM 200
           E +LK+AALWD+VKD+LHK+A +LSGGQQQRLCIARAIAVKPE+LLMDEP SALDPI+T+
Sbjct: 126 EWALKKAALWDEVKDELHKNALSLSGGQQQRLCIARAIAVKPEVLLMDEPTSALDPISTL 185

Query: 201 QLEETMFELKKNYTIIIVTHNMQQAARASDYTAFFYLGDLIEYDKTNNIFQNAKCQSTSD 260
           ++EE M EL K+YTI++VTHNMQQA+R SDYTAFF +G LIE+ +T  IF N + + T D
Sbjct: 186 KIEELMVELAKDYTIVVVTHNMQQASRVSDYTAFFLMGKLIEFGETEQIFLNPQKKETDD 245

Query: 261 YVSGRFG 267
           Y+SGRFG
Sbjct: 246 YISGRFG 252
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1681> which encodes the amino acid sequence <SEQ ID 1682>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2796(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 242/267 (90%), Positives = 258/267 (95%)

Query: 1   MAEYNWDERHIITFPEENSALTTKDLHVYYGEKEAIKGIDMQFEKNKITALIGPSGCGKS 60
            M EYNW+ERHIITFPEE  AL TKDLHVYYG KEAIKGIDMQFEK+KITALIGPSGCGKS
Sbjct: 1   MTEYNWNERHIITFPEETLALATKDLHVYYGAKEAIKGIDMQFEKHKITALIGPSGCGKS 60

Query: 61  TYLRSLNRMNDTIDIARVTGQIMYEGIDVNAQDINVYEMRKHIGMVFQRPNPFAKSIYKN 120
            TYLRSLNRMNDTIDIARVTG+I+Y+GIDVN +D+NVYE+RKH+GMVFQRPNPFAKSIYKN
Sbjct: 61  TYLRSLNRMNDTIDIARVTGEILYQGIDVNRKDMNVYEIRKHLGMVFQRPNPFAKSIYKN 120

Query: 121 ITFAYERAGVKDKKFLDEVVETSLKQAALWDQVKDDLHKSAFTLSGGQQQRLCIARAIAV 180
            ITFA+ERAGVKDKK LDE+VETSLKQAALWDQVKDDLHKSAFTLSGGQQQRLCIARAI+V
Sbjct: 121 ITFAHERAGVKDKKVLDEIVETSLKQAALWDQVKDDLHKSAFTLSGGQQQRLCIARAISV 180

Query: 181 KPEILLMDEPASALDPIATMQLEETMFELKKNYTIIIVTHNMQQAARASDYTAFFYLGDL 240
            KP+ILLMDEPASALDPIATMQLEETMFELKKNYTIIIVTHNMQQAARASDYTAFFYLG+L
Sbjct: 181 KPDILLMDEPASALDPIATMQLEETMFELKKNYTIIIVTHNMQQAARASDYTAFFYLGNL 240

Query: 241 IEYDKTNNIFQNAKCQSTSDYVSGRFG 267
            IEYDKT NIFQNA+CQST+DYVSG FG
Sbjct: 241 IEYDKTRNIFQNAQCQSTNDYVSGHFG 267
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 672

A DNA sequence (GBSx0712) was identified in *S.agalactiae* <SEQ ID 2065> which encodes the amino acid sequence <SEQ ID 2066>. This protein is predicted to be phosphate ABC transporter, ATP-binding protein (pstB-1). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3806(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9815> which encodes amino acid sequence <SEQ ID 9816> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14426 GB:Z99116 alternate gene name: yzmE~similar to
          phosphate ABC transporter (ATP-binding protein)
          [Bacillus subtilis]
 Identities = 148/248 (59%), Positives = 189/248 (75%)

Query: 5   ILQVSDLSVYYNKKKALKEVSMDFYPNEITALIGPSGSGKSTLLRAINRMGDLNPEVTLT 64
            +L+V DLS+YY  K+A+  V+MD   N +TALIGPSG GKST LR INRM DL P
Sbjct: 22  VLEVKDLSIYYGNKQAVHHVNMDIEKNAVTALIGPSGCGKSTFLRNINRMNDLIPSARAE 81

Query: 65  GAVMYNGHNVYSPRTDTVELRKEIGMVFQQPNPFPMSVFENVVYGLRLKGIKDKATLDEA 124
```

```
           G ++Y G N+     + V LR+EIGMVFQ+PNPFP S++ N+ + L+  G ++KA LDE
Sbjct: 82  GEILYEGLNILGGNINVVSLRREIGMVFQKPNPFPKSIYANITHALKYAGERNKAVLDEI 141

Query: 125 VETSLKGASIWDEVKDRLHDSALGLSGGQQQRVCIARTLATKPKIILLDEPTSALDPISA 184
           VE SL  A++WDEVKDRLH SAL LSGGQQQR+CIARTLA KP ++LLDEP SALDPIS
Sbjct: 142 VEESLTKAALWDEVKDRLHSSALSLSGGQQQRLCIARTLAMKPAVLLLDEPASALDPISN 201

Query: 185 GKIEETLHGLKDQYTMLLVTRSMQQASRISDRTGFFLDGNLIEYGNTKEMFMNPKHKETE 244
            KIEE + GLK +Y++++VT +MQQA R+SDRT FFL+G L+EYG T+++F +PK ++TE
Sbjct: 202 AKIEELITGLKREYSIIIVTHNMQQALRVSDRTAFFLNGELVEYGQTEQIFTSPKKQKTE 261

Query: 245 DYITGKFG 252
           DYI GKFG
Sbjct: 262 DYINGKFG 269
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2067> which encodes the amino acid sequence <SEQ ID 2068>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3590(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 208/252 (82%), Positives = 235/252 (92%)

Query: 1   MTQPILQVSDLSVYYNKKKALKEVSMDFYPNEITALIGPSGSGKSTLLRAINRMGDLNPE 60
           MT+PILQ+ DLSVYYN+KK LK+VS+D YPNEITALIGPSGSGKSTLLR+INRM DLNPE
Sbjct: 2   MTEPILQIRDLSVYYNQKKTLKDVSLDLYPNEITALIGPSGSGKSTLLRSINRMNDLNPE 61

Query: 61  VTLTGAVMYNGHNVYSPRTDTVELRKEIGMVFQQPNPFPMSVFENVVYGLRLKGIKDKAT 120
           VT+TG+++YNGHN+YSPRTDTV+LRKEIGMVFQQPNPFPMS++ENVVYGLRLKGI+DK+
Sbjct: 62  VTITGSIVYNGHNIYSPRTDTVDLRKEIGMVFQQPNPFPMSIYENVVYGLRLKGIRDKSI 121

Query: 121 LDEAVETSLKGASIWDEVKDRLHDSALGLSGGQQQRVCIARTLATKPKIILLDEPTSALD 180
           LD AVE+SLKGASIW+EVKDRLHDSA+GLSGGQQQRVCIAR LAT P+IILLDEPTSALD
Sbjct: 122 LDHAVESSLKGASIWNEVKDRLHDSAVGLSGGQQQRVCIARVLATSPRIILLDEPTSALD 181

Query: 181 PISAGKIEETLHGLKDQYTMLLVTRSMQQASRISDRTGFFLDGNLIEYGNTKEMFMNPKH 240
           PISAGKIEETL  LK  YT+ +VTRSMQQASR+SDRTGFFL+G+L+E G TK MFMNPK
Sbjct: 182 PISAGKIEETLLLLKKDYTLAIVTRSMQQASRLSDRTGFFLEGDLLECGPTKAMFMNPKR 241

Query: 241 KETEDYITGKFG 252
           KETEDYI+GKFG
Sbjct: 242 KETEDYISGKFG 253
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 673

A DNA sequence (GBSx0713) was identified in *S.agalactiae* <SEQ ID 2069> which encodes the amino acid sequence <SEQ ID 2070>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1937(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD22042 GB:AF118229 PhoU [Streptococcus pneumoniae]
 Identities = 75/216 (34%), Positives = 126/216 (57%), Gaps = 1/216 (0%)

Query: 2   LRSKFDEELDKLHNQFYAMGIEAIGQIKKTVRAFVSHDRELAKEVIEDDVTLNNFETKLE 61
           +R++FD EL +L   F  +G   +    K + A  S D+E+A+ +I  D  +N  ++ +E
Sbjct: 1   MRNQFDLELHELEQSFLGLGQLVLETASKALLALASKDKEMAELIINKDHAINQGQSAIE 60

Query: 62  KKSLEIIALQQPVSQDLRTVITVLKATSDVERMGDHAAAVAKATIRMKGEERIPAVELEI 121
                ++ALQQP   DLR VI+++ + SD+ERMGDH A +AKA +++K E ++   E ++
Sbjct: 61  LTCARLLALQQPQVSDLRFVISIMSSCSDLERMGDHMAGIAKAVLQLK-ENQLAPDEEQL 119

Query: 122 NNMGKAVKNMLEEALTAYINGDDEKAYEVAAMDEIVDDYFRDIQKMVVETIQKHPDVAFA 181
           + MGK   +ML + L A+      KA  +A  DE +D Y+  + K ++  ++
Sbjct: 120 HQMGKLSLSMLADLLVAFPLHQASKAISIAQKDEQIDQYYYALSKEIIGLMKDQETSIPN 179

Query: 182 AKEYFQVLMHLERIGDYGKNICEWIVYLKTGKIIEL 217
             +Y  ++ HLER  DY  NICE +VYL+TG++++L
Sbjct: 180 GTQYLYIIGHLERFADYIANICERLVYLETGELVDL 215
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1677> which encodes the amino acid sequence <SEQ ID 1678>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2229(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 174/217 (80%), Positives = 194/217 (89%)

Query: 1   MLRSKFDEELDKLHNQFYAMGIEAIGQIKKTVRAFVSHDRELAKEVIEDDVTLNNFETKL 60
           MLR+KF+EELDKLHNQFY+MG+E + QI KTVRAFVSHDRELAKEVIE+D T+NNFETKL
Sbjct: 1   MLRTKFEEELDKLHNQFYSMGMEVLAQINKTVRAFVSHDRELAKEVIEEDDTINNFETKL 60

Query: 61  EKKSLEIIALQQPVSQDLRTVITVLKATSDVERMGDHAAAVAKATIRMKGEERIPAVELE 120
           EKKSLEIIALQQPVS DLR VITVLKA+SD+ERMGDHAA++AKATIRMKGEERIP VE +
Sbjct: 61  EKKSLEIIALQQPVSNDLRMVITVLKASSDIERMGDHAASIAKATIRMKGEERIPVVEEQ 120

Query: 121 INNMGKAVKNMLEEALTAYINGDDEKAYEVAAMDEIVDDYFRDIQKMVVETIQKHPDVAF 180
           IN MGKAVK M+EEAL AYIN DD KAYE+AA DEI+D YFR+IQ + VE I+K PD  F
Sbjct: 121 INLMGKAVKQMVEEALNAYINADDTKAYEIAASDEIIDQYFRNIQTLAVEEIRKSPDAVF 180

Query: 181 AAKEYFQVLMHLERIGDYGKNICEWIVYLKTGKIIEL 217
           A KEYFQVLM+LERIGDY +NICEWIVYLKTGKIIEL
Sbjct: 181 AGKEYFQVLMYLERIGDYARNICEWIVYLKTGKIIEL 217
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 674

A DNA sequence (GBSx0714) was identified in *S.agalactiae* <SEQ ID 2071> which encodes the amino acid sequence <SEQ ID 2072>. This protein is predicted to be aminopeptidase N. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2845(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB50785 GB:AJ007700 aminopeptidase N [Streptococcus thermophilus]
 Identities = 556/847 (65%), Positives = 673/847 (78%), Gaps = 4/847 (0%)

Query: 3   TVEHFVTKFVPENYNLFLDINRQTKTFSGNVAVSGEALDNNISFHQKGLTIKSVLLDNQP 62
            +V  F+  F+PENYNLFLDINR  KTF+GNVA++GEA+DN+IS HQK LTI SVLLDN+
Sbjct: 4   SVARFIESFIPENYNLFLDINRSEKTFTGNVAITGEAIDNHISLHQKDLTINSVLLDNES 63

Query: 63  LDFQLDEDNEAMHIQLHETGSMVLVFEFSGHITDNMTGMYPSYYTVNGIKKEVISTQFES 122
            L+FQ+D+ NEA HI+L ETG + +  EFSG ITDNMTG+YPSYYT NG KKE+ISTQFES
Sbjct: 64  LNFQMDDANEAFHIELPETGVLTIFIEFSGRITDNMTGIYPSYYTYNGEKKEIISTQFES 123

Query: 123 HFAREVFPSIDEPEAKATFDLSLKFDQKEGEIALSNMPEINAEQRQETGLWTFDTTPKMS 182
            HFARE FP +DEPEAKATFDLSLKFD +EG+ ALSNMPEIN+  R+ETG+WTF+TTP+MS
Sbjct: 124 HFAREAFPCVDEPEAKATFDLSLKFDAEEGDTALSNMPEINSHLREETGVWTFETTPRMS 183

Query: 183 SYLLAFALGELHGKTTHTKNGTLVGSYATKAHQLNELDFSLDIVVRVIEFYEDYFGVRYP 242
            +YLLAF  G LHGKT  TKNGT VG +AT A   N +DF+LDI VRVIEFYEDYF V+YP
Sbjct: 184 TYLLAFGFGALHGKTAKTKNGTEVGVFATVAQAENSVDFALDIAVRVIEFYEDYFQVKYP 243

Query: 243 IPQSLHVALPDFSAGAMENWGLVTYREVYLLVDENSSVSSRQQVALVVAHEIAHQWFGNL 302
            IP S H+ALPD SAGAMENWGLVTYREVYLLVDENSS +SRQQVALVVAHE+AHQWFGNL
Sbjct: 244 IPLSYHLALPDLSAGAMENWGLVTYREVYLLVDENSSAASRQQVALVVAHELAHQWFGNL 303

Query: 303 VTMKWWDDLWLNESFANMMEYVSIDYIEPKLNIFEDFQTG-GLPLALKRDATDGVQSVHV 361
            VTMKWWDDLWLNESFANMMEYVS++ IEP  NIFE F    G+P AL+RDATDGVQSVH+
Sbjct: 304 VTMKWWDDLWLNESFANMMEYVSVNAIEPSWNIFEGFPNKLGVPNALQRDATDGVQSVHM 363

Query: 362 EVNHPDEINTLFDPAIVYAKGSRLMHMLRRWLGDTDFAAGLKIYFEKHQYQNTIGRDLWN 421
            EVNHPDEINTLFD AIVYAKGSRLMHMLRRWLGD  FA GLK YFEKHQY NT+GRDLWN
Sbjct: 364 EVNHPDEINTLFDSAIVYAKGSRLMHMLRRWLGDEAFAKGLKAYFEKHQYNNTVGRDLWN 423

Query: 422 ALSQTSGKDVAAFMDSWLEQPGYPVMAAKIEEDELILTQKQFFIGEHEDKSRLWQIPLNS 481
            ALS+ SGKDV++FMD+WLEQPGYPV++A++ +D LIL+QKQFFIGEHEDK RLW+IPLN+
Sbjct: 424 ALSEASGKDVSSFMDTWLEQPGYPVVSAEVVDDTLILSQKQFFIGEHEDKGRLWEIPLNT 483

Query: 482 NWEGIPEILTEETVVIPNFSQLAEKNKENGALRFNTENTAHYITNYQGQLLEHIISDLPL 541
            NW G+P+ L+EE + IPN+SQLA +N  NG LR NT NTAHYIT+YQGQLL++I+ D
Sbjct: 484 NWNGLPDTLSEERIEIPNYSQLATEN--NGVLRLNTANTAHYITDYQGQLLDNILEDFAN 541

Query: 542 MDNISKLQIVQERHLLAESGMISYSSLIPLVSLLSQETSYLVNSAIKSVIDGLSLFVQED 601
            +D +SKLQI+QER LLAESG ISY+SL+ L+ L+ +E S+L++ A   ++ GL  F+ ED
Sbjct: 542 LDTVSKLQILQERRLLAESGRISYASLVGLLDLVEKEESFLISQAKSQILAGLKRFIDED 601

Query: 602 SQDEFDFKEFVNKLSAFNFNRLGFEKREGEGDDSEMVRHLSLSLALYSDNEHAIEEAHHI 661
            ++ E  +K  V++    +F RLGF+ +EGE D+ EMVR  +LS  + +D +  +  A ++
Sbjct: 602 TEAEVHYKALVSRQFQNDFERLGFDAKEGESDEDEMVRQTALSYLIEADYQPTVLAAANV 661

Query: 662 FKAHENNIAAIPAAIRLLVLTNEMKHFESKELSHLLLETYSTTTDGNFKRQLASALSHTT 721
            F+AH+ NI +IPA+IR LVL N+MK   S  L    +  Y  T D NF+RQL  ALS+
Sbjct: 662 FQAHKENIESIPASIRGLVLINQMKQENSLSLVEEYINAYVATNDSNFRRQLTQALSYLK 721

Query: 722 DSKTLKKLLSDWKNKDIVKPQDLAMSWYATFLKNSFTQESVWEWAQENWEWIKATLGGDM 781
            + + L  +L   K+K++VKPQDL + WY  FL  SF QE+VW+WA+ENWEWIKA LGGDM
Sbjct: 722 NQEGLDYVLGQLKDKNVVKPQDLYL-WYMNFLSKSFAQETVWDWAKENWEWIKAALGGDM 780

Query: 782 SFDKFVIYPSSSFKTEERLEQYKNFFEPQLSDMAISRNISMGIKEISARVLLITKQKEEV 841
            SFD FV  P+  FK +ERL+QY  FFEPQ SD A+ RNI MGIK I+ARV LI K+K  V
Sbjct: 781 SFDSFVNIPAGIFKNQERLDQYIAFFEPQTSDKALERNILMGIKTIAARVDLIEKEKAAV 840

Query: 842 INTIKKY 848
```

```
           + +K Y
Sbjct: 841 ESALKDY 847
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2073> which encodes the amino acid sequence <SEQ ID 2074>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1098(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 576/848 (67%), Positives = 692/848 (80%), Gaps = 3/848 (0%)

Query: 1   MKTVEHFVTKFVPENYNLFLDINRQTKTFSGNVAVSGEALDNNISFHQKGLTIKSVLLDN 60
           MKTVEH +  FVPENYN+FLDINRQTKTF+GNVA++GEALDN+++FHQK L IKS+LLDN
Sbjct: 21  MKTVEHLIETFVPENYNIFLDINRQTKTFTGNVAINGEALDNHVAFHQKDLDIKSILLDN 80

Query: 61  QPLDFQLDEDNEAMHIQLHETGSMVLVFEFSGHITDNMTGMYPSYYTVNGIKKEVISTQF 120
           + + +Q+D DNE + ++L ETG M LV EFSG ITDNMTG+YPSYYT NG KKEVISTQF
Sbjct: 81  EAVIYQVDNDNEVVRVELPETGMMTLVIEFSGSITDNMTGIYPSYYTKNGEKKEVISTQF 140

Query: 121 ESHFAREVFPSIDEPEAKATFDLSLKFDQKEGEIALSNMPEINAEQRQETGLWTFDTTPK 180
           ESHFARE FP IDEP+AKATFDLSL FDQ+ GEIALSNMPE+N ++R+ETGLWTFDTT +
Sbjct: 141 ESHFAREAFPCIDEPQAKATFDLSLTFDQEIGEIALSNMPEVNIDRREETGLWTFDTTLR 200

Query: 181 MSSYLLAFALGELHGKTTHTKNGTLVGSYATKAHQLNELDFSLDIVVRVIEFYEDYFGVR 240
           MSSYLLAFALGELHGKT  +K GT VG YAT AH L+ LDFSLDI VRVI FYEDYFGV
Sbjct: 201 MSSYLLAFALGELHGKTVESKKGTTVGVYATTAHPLSSLDFSLDIAVRVINFYEDYFGVH 260

Query: 241 YPIPQSLHVALPDFSAGAMENWGLVTYREVYLLVDENSSVSSRQQVALVVAHEIAHQWFG 300
           YPIPQSL++ALPDFS+GAMENWGL+TYRE+YLLVDENS+V SRQQVALV+AHEIAHQWFG
Sbjct: 261 YPIPQSLNIALPDFSSGAMENWGLITYREIYLLVDENSTVQSRQQVALVIAHEIAHQWFG 320

Query: 301 NLVTMKWWDDLWLNESFANMMEYVSIDYIEPKLNIFEDFQTGGLPLALKRDATDGVQSVH 360
           NLVTMKWWDDLWLNESFANMMEYVSI+ IEP   I EDFQTGG+PLALKRDATDGVQSVH
Sbjct: 321 NLVTMKWWDDLWLNESFANMMEYVSIEAIEPSWKIIEDFQTGGIPLALKRDATDGVQSVH 380

Query: 361 VEVNHPDEINTLFDPAIVYAKGSRLMHMLRRWLGDTDFAAGLKIYFEKHQYQNTIGRDLW 420
           VEVNHPDEINTLFDPAIVYAKGSRLMHMLRR++GD DFA GL  YFEK+QY+NT+GRDLW
Sbjct: 381 VEVNHPDEINTLFDPAIVYAKGSRLMHMLRRFIGDRDFAIGLHHYFEKYQYRNTVGRDLW 440

Query: 421 NALSQTSGKDVAAFMDSWLEQPGYPVMAAKIEEDELILTQKQFFIGEHEDKSRLWQIPLN 480
           N LS TSGKDVAAFMD+WLEQPGYPV+ A++E D+LIL+QKQFFIG+ E+K RLW IPLN
Sbjct: 441 NILSDTSGKDVAAFMDAWLEQPGYPVLTARLENDQLILSQKQFFIGKGEEKGRLWPIPLN 500

Query: 481 SNWEGIPEILTEETVVIPNFSQLAEKNKENGALRFNTENTAHYITNYQGQLLEHIISDLP 540
           +NW G+PE LTE  +VIPNFSQLA +N+  GALRFN +NTAHYIT+YQG LL+ ++++L
Sbjct: 501 TNWHGLPETLTEAEMVIPNFSQLAAENE--GALRFNIDNTAHYITDYQGSLLDALVTELA 558

Query: 541 LMDNISKLQIVQERHLLAESGMISYSSLIPLVSLLSQETSYLVNSAIKSVIDGLSLFVQE 600
            +DN S LQ++QER LLA+SG+ISY+ L+ L++ L    SY+V  A++ V+ GL  F+ E
Sbjct: 559 QLDNTSALQVIQERRLLADSGLISYAELVDLIAQLDDSKSYMVAEAVQQVVSGLKRFIDE 618

Query: 601 DSQDEFDFKEFVNKLSAFNFNRLGFEKREGEGDDSEMVRHLSLSLALYSDNEHAIEEAHH 660
            S  E  F   V  +   +FN+ GFEK+  E D+ EMVR ++L     ++N+  I+
Sbjct: 619 GSLAEKSFNRLVTTIYQEDFNQHGFEKKADESDEDEMVRQVALGRLWLAENQTIIDGLRT 678

Query: 661 IFKAHENNIAAIPAAIRLLVLTNEMKHFESKELSHLLLETYSTTTDGNFKRQLASALSHT 720
           IF+A++NNIA+IPAA+R LVL N+MK+FE+  L  +  ETY  TTD N +  L  A S T
Sbjct: 679 IFEAYQNNIASIPAAVRRLVLANQMKYFETDSLVDIYFETYVATTDNNLRSDLTVAFSQT 738

Query: 721 TDSKTLKKLLSDWKNKDIVKPQDLAMSWYATFLKNSFTQESVWEWAQENWEWIKATLGGD 780
```

```
               T++++L   K+KDI+KPQDL+  WY   L  SFTQ+ +WEWA+ENW+WIK+ LGGD
Sbjct: 739 KQPTTIRRILVSLKDKDIIKPQDLSY-WYNALLGQSFTQDIIWEWARENWDWIKSALGGD 797

Query: 781 MSFDKFVIYPSSSFKTEERLEQYKNFFEPQLSDMAISRNISMGIKEISARVLLITKQKEE 840
            MSFDKFVIYP+S+FKT + L +YK+FFEP+L DMAISRNI+MGI EI ARV LITK+KE
Sbjct: 798 MSFDKFVIYPASNFKTPKHLAEYKSFFEPKLDDMAISRNITMGINEIEARVALITKEKEA 857

Query: 841 VINTIKKY 848
            VI  +  Y
Sbjct: 858 VIAALSHY 865
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 675

A DNA sequence (GBSx0715) was identified in *S.agalactiae* <SEQ ID 2075> which encodes the amino acid sequence <SEQ ID 2076>. This protein is predicted to be response regulator (trcR). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2741(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA54465 GB:X77249 response regulator [Streptococcus pneumoniae]
 Identities = 198/224 (88%), Positives = 213/224 (94%)

Query: 1   MIKILLIEDDLSLSNSVFDFLDDFADVMQIFDGEEGLYEAESGVYDLILLDLMLPEKNGF 60
           MIKILL+EDDL LSNSVFDFLDDFADVMQ+FDGEEGLYEAESGVYDLILLDLMLPEKNGF
Sbjct: 1   MIKILLVEDDLGLSNSVFDFLDDFADVMQVFDGEEGLYEAESGVYDLILLDLMLPEKNGF 60

Query: 61  QVLKELREKGITTPVLIMTAKESIDDKGQGFDLGADDYLTKPFYLEELKMRIQALLKRSG 120
           QVLKELREKGITTPVLIMTAKES+DDKG GF+LGADDYLTKPFYLEELKMRIQALLKRSG
Sbjct: 61  QVLKELREKGITTPVLIMTAKESLDDKGHGFELGADDYLTKPFYLEELKMRIQALLKRSG 120

Query: 121 KFNDNSLIYGDIRVDMSTNSTFVNQTEVELLGKEFDLLVYFLQNQNVILPKSQIFDRIWG 180
           KFN+N+L YG+I V++STN+  V  T VELLGKEFDLLVYFLQNQNVILPK+QIFDR+WG
Sbjct: 121 KFNENTLTYGNIVVNLSTNTVKVEDTPVELLGKEFDLLVYFLQNQNVILPKTQIFDRLWG 180

Query: 181 FDSDTTISVVEVYVSKVRKKLKGTLFSENLQTLRSVGYILKHVE 224
           FDSDTTISVVEVYVSKVRKKLKGT F+ENLQTLRSVGY+LK V+
Sbjct: 181 FDSDTTISVVEVYVSKVRKKLKGTTFAENLQTLRSVGYLLKDVQ 224
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2077> which encodes the amino acid sequence <SEQ ID 2078>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2689(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 180/224 (80%), Positives = 200/224 (88%)
```

```
Query: 1   MIKILLIEDDLSLSNSVFDFLDDFADVMQIFDGEEGLYEAESGVYDLILLDLMLPEKNGF 60
           MIKILL+EDDLSLSNS+FDFLDDFADVMQ+FDG+EGLYEAESG+YDLILLDLMLPEKNGF
Sbjct: 1   MIKILLVEDDLSLSNSIFDFLDDFADVMQVFDGDEGLYEAESGIYDLILLDLMLPEKNGF 60

Query: 61  QVLKELREKGITTPVLIMTAKESIDDKGQGFDLGADDYLTKPFYLEELKMRIQALLKRSG 120
           QVLKELREK I  PVLIMTAKE +DDKG GF+LGADDYLTKPFYLEELKMRIQALLKR+G
Sbjct: 61  QVLKELREKDIKIPVLIMTAKEGLDDKGHGFELGADDYLTKPFYLEELKMRIQALLKRTG 120

Query: 121 KFNDNSLIYGDIRVDMSTNSTFVNQTEVELLGKEFDLLVYFLQNQNVILPKSQIFDRIWG 180
           KF D ++ +G++ VD++     V    VELLGKEFDLLVY LQNQNVILPK+QIFDR+WG
Sbjct: 121 KFADKNISFGNLVVDLARKEVKVEGKVVELLGKEFDLLVYLLQNQNVILPKTQIFDRLWG 180

Query: 181 FDSDTTISVVEVYVSKVRKKLKGTLFSENLQTLRSVGYILKHVE 224
           FDSDTTISVVEVY+SK+RKKLKGT F   LQTLRSVGYILK+ E
Sbjct: 181 FDSDTTISVVEVYISKIRKKLKGTCFVNRLQTLRSVGYILKNNE 224
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 676

A DNA sequence (GBSx0716) was identified in *S.agalactiae* <SEQ ID 2079> which encodes the amino acid sequence <SEQ ID 2080>. This protein is predicted to be histidine kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.18   Transmembrane   22 -  38 (  17 -  46)
   INTEGRAL    Likelihood = -4.94   Transmembrane  182 - 198 ( 178 - 201)

----- Final Results -----
              bacterial membrane --- Certainty=0.4673(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA54466 GB:X77249 histidine kinase [Streptococcus pneumoniae]
 Identities = 218/420 (51%), Positives = 305/420 (71%), Gaps = 4/420 (0%)

Query: 17  SHFIHFFTVFSGIFLVMTVIILQVMRYGVYSSVDSSLKYISTHPKNYINMVMSRTAAY-- 74
           S+FI  F VF+ IF  MT+IILQVM   +Y+SVD  L  +S +P+  I + ++R
Sbjct: 15  SYFIRNFGVFTLIFSTMTLIILQVMHSSLYTSVDDKLHGLSENPQAVIQLAINRATEEIK 74

Query: 75  -LDNSNIASVKLKPGGQTVANTDIILFTSEEEVINYFDAFSNYQFLKPNKKNLGGISELT 133
            L+N+    + K++      +NT++ILF  +   +   + F     +K  KK LG I ++
Sbjct: 75  DLENARADASKVEIKPNVSSNTEVILFDKDFTQLLSGNRFLGLDKIKLEKKELGHIYQIQ 134

Query: 134 LTNIFGQDETYHAVTVKVN-NPAYPNVTYMTAIVNIDQLVNAKERYEKIIIFVMTTFWII 192
           + N +GQ+E Y  + ++ N +    N+ Y   ++N  QL  A +++E++I+ VM +FWI+
Sbjct: 135 VFNSYGQEEIYRVILMETNISSVSTNIKYAAVLINTSQLEQASQKHEQLIVVVMASFWIL 194

Query: 193 SIGASIYLAKWAQKPIIENYERQKAFVENASHELRTPLAVLQNRLETLFRKPNATILENS 252
           S+ AS+YLA+ + +P++E+ ++Q++FVENASHELRTPLAVLQNRLETLFRKP ATI++ S
Sbjct: 195 SLLASLYLARVSVRPLLESMQKQQSFVENASHELRTPLAVLQNRLETLFRKPEATIMDVS 254

Query: 253 ENIASSLDEVRNMRILTTNLLNLARRDDGIKPELAVIKPTLFDSIFENYDLITQENGKNF 312
           E+IASSL+EVRNMR LTT+LLNLARRDDGIKPELA +  + F++ F NY++I  EN + F
Sbjct: 255 ESIASSLEEVRNMRFLTTSLLNLARRDDGIKPELAEVPTSFFNTTFTNYEMIASENNRVF 314

Query: 313 TGHNMIQDSFKTDKTLLKQLMTILFDNAIKYTDNDGSIDFTISETDKYLFLEIADNGPGI 372
              N I  +  TD+ LLKQLMTILFDNA+KYT+ DG IDF IS TD+ L+L ++DNG GI
Sbjct: 315 RFENRIHRTIVTDQLLLKQLMTILFDNAVKYTEEDGEIDFLISATDRNLYLLVSDNGIGI 374

Query: 373 SEEDKVRIFDRFYRVDKARTRQQGGFGLGLSLAQQIVNSLRGNITVIDNKPRGSIFKIKL 432
```

```
           S EDK +IFDRFYRVDKARTRQ+GGFGLGLSLA+QIV++L+G +TV DNKP+G+IF++K+
Sbjct: 375 STEDKKKIFDRFYRVDKARTRQKGGFGLGLSLAKQIVDALKGTVTVKDNKPKGTIFEVKI 434
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2081> which encodes the amino acid sequence <SEQ ID 2082>. Analysis of this protein sequence reveals the following:

```
     Possible site: 57
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.09   Transmembrane   19 -  35 (  14 -  44)
   INTEGRAL    Likelihood =-10.24   Transmembrane  185 - 201 ( 182 - 206)

----- Final Results -----
              bacterial membrane --- Certainty=0.5437(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA54466 GB:X77249 histidine kinase [Streptococcus pneumoniae]
 Identities = 223/436 (51%), Positives = 313/436 (71%), Gaps = 5/436 (1%)

Query: 2   NKLKKEILSDNYNHFFHFFAVFTGIFVIMTIIILQIMRFGVYSSVDSSLVSVSNNASSYA 61
           +KLKK   +D++++F   F VFT IF  MT+IILQ+M   +Y+SVD  L  +S N  +
Sbjct: 3   SKLKKTWYADDFSYFIRNFGVFTLIFSTMTLIILQVMHSSLYTSVDDKLHGLSENPQAVI 62

Query: 62  NRTMARISSFYFDTENNIIKALPDSDSSKLLGTPAANTDIILFSANGTILNAFDAFSNYQ 121
              + R +    D EN    A  D+   ++    ++NT++ILF  + T L + + F
Sbjct: 63  QLAINRATEEIKDLEN----ARADASKVEIKPNVSSNTEVILFDKDFTQLLSGNRFLGLD 118

Query: 122 NFHLDKRRLGSIETTSLMNFYGQEEKYHTITVGVHIKNYPA-VAYMMAVVNVEQLDRANE 180
              L+K+ LG I    + N YGQEE Y  I +  +I +    + Y   ++N  QL++A++
Sbjct: 119 KIKLEKKELGHIYQIQVFNSYGQEEIYRVILMETNISSVSTNIKYAAVLINTSQLEQASQ 178

Query: 181 RYERIIIIVMSVFWLISILASIYLAKWSRKPILESYEKQKMFVENASHELRTPLAVLQNR 240
           ++E++I++VM+ FW++S+LAS+YLA+ S +P+LES +KQ+ FVENASHELRTPLAVLQNR
Sbjct: 179 KHEQLIVVVMASFWILSLLASLYLARVSVRPLLESMQKQQSFVENASHELRTPLAVLQNR 238

Query: 241 LESLFRKPNETILENSEHLASSLDEVRNMRILTTNLLNLARRDDGINPQWTHLDTDFFNA 300
           LE+LFRKP  TI++ SE +ASSL+EVRNMR LTT+LLNLARRDDGI P+   + T FFN
Sbjct: 239 LETLFRKPEATIMDVSESIASSLEEVRNMRFLTTSLLNLARRDDGIKPELAEVPTSFFNT 298

Query: 301 IFENYELVAKEYGKIFYFQNQVNRSLRMDKALLKQLITILFDNAIKYTDKNGIIEIIVKT 360
            F NYE++A E  ++F F+N+++R++  D+ LLKQL+TILFDNA+KYT+++G I+ ++
Sbjct: 299 TFTNYEMIASENNRVFRFENRIHRTIVTDQLLLKQLMTILFDNAVKYTEEDGEIDFLISA 358

Query: 361 TDKNLLISVIDNGPGITDEEKKKIFDRFYRVDKARTRQTGGFGLGLALAQQIVMSLKGNI 420
           TD+NL + V DNG GI+ E+KKKIFDRFYRVDKARTRQ GGFGLGL+LA+QIV +LKG +
Sbjct: 359 TDRNLYLLVSDNGIGISTEDKKKIFDRFYRVDKARTRQKGGFGLGLSLAKQIVDALKGTV 418

Query: 421 TVKDNDPKGSIFEVKL 436
           TVKDN PKG+IFEVK+
Sbjct: 419 TVKDNKPKGTIFEVKI 434
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 265/436 (60%), Positives = 334/436 (75%), Gaps = 10/436 (2%)

Query: 7   ISKFKKNV-SDS--HFIHFFTVFSGIFLVMTVIILQVMRYGVYSSVDSSLKYISTHPKNY 63
           ++K KK + SD+  HF HFF VF+GIF++MT+IILQ+MR+GVYSSVDSSL  +S +  +Y
Sbjct: 1   MNKLKKEILSDNYNHFFHFFAVFTGIFVIMTIIILQIMRFGVYSSVDSSLVSVSNNASSY 60

Query: 64  INMVMSRTAAYLDNSNIASVKLKPG-------GQTVANTDIILFTSEEEVINYFDAFSNY 116
            N  M+R +++  ++    +K  P        G   ANTDIILF++   ++N FDAFSNY
Sbjct: 61  ANRTMARISSFYFDTENNIIKALPDSDSSKLLGTPAANTDIILFSANGTILNAFDAFSNY 120

Query: 117 QFLKPNKKNLGGISELTLTNIFGQDETYHAVTVKVNNPAYPNVTYMTAIVNIDQLVNAKE 176
           Q    +K+ LG I  +L  N +GQ+E YH +TV V+    YP V YM A+VN++QL  A E
```

```
Sbjct: 121 QNFHLDKRRLGSIETTSLMNFYGQEEKYHTITVGVHIKNYPAVAYMMAVVNVEQLDRANE 180

Query: 177 RYEKIIIFVMTTFWIISIGASIYLAKWAQKPIIENYERQKAFVENASHELRTPLAVLQNR 236
            RYE+III VM+ FW+ISI ASIYLAKW++KPI+E+YE+QK FVENASHELRTPLAVLQNR
Sbjct: 181 RYERIIIIVMSVFWLISILASIYLAKWSRKPILESYEKQKMFVENASHELRTPLAVLQNR 240

Query: 237 LETLFRKPNATILENSENIASSLDEVRNMRILTTNLLNLARRDDGIKPELAVIKPTLFDS 296
            LE+LFRKPN TILENSE++ASSLDEVRNMRILTTNLLNLARRDDGI P+   +    F++
Sbjct: 241 LESLFRKPNETILENSEHLASSLDEVRNMRILTTNLLNLARRDDGINPQWTHLDTDFFNA 300

Query: 297 IFENYDLITQENGKNFTGHNMIQDSFKTDKTLLKQLMTILFDNAIKYTDNDGSIDFTISE 356
            IFENY+L+ +E GK F   N +  S + DK LLKQL+TILFDNAIKYTD +G I+   +
Sbjct: 301 IFENYELVAKEYGKIFYFQNQVNRSLRMDKALLKQLITILFDNAIKYTDKNGIIEIIVKT 360

Query: 357 TDKYLFLEIADNGPGISEEDKVRIFDRFYRVDKARTRQQGGFGLGLSLAQQIVNSLRGNI 416
            TDK L + + DNGPGI++E+K +IFDRFYRVDKARTRQ GGFGLGL+LAQQIV SL+GNI
Sbjct: 361 TDKNLLISVIDNGPGITDEEKKKIFDRFYRVDKARTRQTGGFGLGLALAQQIVMSLKGNI 420

Query: 417 TVIDNKPRGSIFKIKL 432
            TV DN P+GSIF++KL
Sbjct: 421 TVKDNDPKGSIFEVKL 436
```

SEQ ID 2080 (GBS339d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 9; MW 73kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 185 (lane 5; MW 73kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 677

A DNA sequence (GBSx0717) was identified in *S.agalactiae* <SEQ ID 2083> which encodes the amino acid sequence <SEQ ID 2084>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1783(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9813> which encodes amino acid sequence <SEQ ID 9814> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB48049 GB:U88582 YlxM [Streptococcus mutans]
 Identities = 95/110 (86%), Positives = 103/110 (93%)

Query: 1    MEIEKTNRMNALFEFYAALLTDKQMNYIELYYADDYSLAEIAEESGVSRQAVYDNIKRTE 60
            MEIEKTNRMNALFEFYAALLTDKQMNYIELYYADDYSLAEIAEE  VSRQAVYDNIKRTE
Sbjct: 1    MEIEKTNRMNALFEFYAALLTDKQMNYIELYYADDYSLAEIAEEFDVSRQAVYDNIKRTE 60

Query: 61   KILEAYEMKLHMYSDYIVRSQIFDDILEKYTDDAFLQEKISILSSIDNRD 110
            KILE YEMKLHMYSDY+VRS+IFD I++KY +D +LQ KISIL++IDNRD
Sbjct: 61   KILEDYEMKLHMYSDYVVRSEIFDAIMKKYPNDPYLQNKISILTTIDNRD 110
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2085> which encodes the amino acid sequence <SEQ ID 2086>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1767(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 95/110 (86%), Positives = 103/110 (93%)

Query: 1   MEIEKTNRMNALFEFYAALLTDKQMNYIELYYADDYSLAEIAEESGVSRQAVYDNIKRTE 60
           MEIEKTNRMNALFEFYAALLTDKQMNYIELYYADDYSLAEIA+E GVSRQAVYDNIKRTE
Sbjct: 4   MEIEKTNRMNALFEFYAALLTDKQMNYIELYYADDYSLAEIADEFGVSRQAVYDNIKRTE 63

Query: 61  KILEAYEMKLHMYSDYIVRSQIFDDILEKYTDDAFLQEKISILSSIDNRD 110
           KILE YEMKLHMYSDY+VRS+IFDD++   Y  D +LQEKISIL+SIDNR+
Sbjct: 64  KILETYEMKLHMYSDYVVRSEIFDDMIAHYPHDEYLQEKISILTSIDNRE 113
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 678

A DNA sequence (GBSx0719) was identified in *S.agalactiae* <SEQ ID 2087> which encodes the amino acid sequence <SEQ ID 2088>. This protein is predicted to be signal recognition particle protein (ffh). Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22   Transmembrane   37 -   53 (  37 -   53)

----- Final Results -----
               bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB48050 GB:U88582 Ffh [Streptococcus mutans]
 Identities = 437/522 (83%), Positives = 484/522 (92%), Gaps = 7/522 (1%)

Query: 1   MAFESLTERLQGVFKNIRGKKKLSEKDVQEVTKEIRLALLEADVALPVVKTFIKHVRERA 60
           MAFESLTERLQGVFKN+RGK+KLSEKDVQEVTKEIRLALLEADVALPVVK FIK VR+RA
Sbjct: 1   MAFESLTERLQGVFKNLRGKRKLSEKDVQEVTKEIRLALLEADVALPVVKEFIKRVRKRA 60

Query: 61  VGHEIIDTLDPTQQIVKIVNEELTDLLGAETSEIEKSPKIPTIIMMVGLQGAGKTTFAGK 120
           VGHE+IDTLDP+QQI+KIVNEELT +LG+ET+EIEKS KIPTIIMMVGLQGAGKTTFAGK
Sbjct: 61  VGHEVIDTLDPSQQIIKIVNEELTAVLGSETAEIEKSSKIPTIIMMVGLQGAGKTTFAGK 120

Query: 121 LANKLIKEDNARPMMIAADIYRPAAIDQLKTLGSQINVPVFDMGTNHSAVEIVTKGLEQA 180
           LANKL+KE+NARP+MIAADIYRPAAIDQLK LG QINVPVFDMGT HSAVEIV++GL QA
Sbjct: 121 LANKLVKEENARPLMIAADIYRPAAIDQLKILGQQINVPVFDMGTEHSAVEIVSQGLAQA 180

Query: 181 RENRNDYVLIDTAGRLQIDATLMQELHDVKAIAQPNEILLVVDSMIGQEAANVAEEFNRQ 240
           +ENRNDYVLIDTAGRLQID  LM EL D+KA+A PNEILLVVDSMIGQEAANVA EFN+Q
Sbjct: 181 KENRNDYVLIDTAGRLQIDEKLMTELRDIKALANPNEILLVVDSMIGQEAANVAREFNQQ 240

Query: 241 LSISGVVLTKIDGDTRGGAALSVREITGKPIKFTGTGEKITDIETFHPDRMASRILGMGD 300
           L ++GV+LTKIDGDTRGGAALSVR+ITGKPIKFTGTGEKITDIETFHPDRM+SRILGMGD
Sbjct: 241 LEVTGVILTKIDGDTRGGAALSVRQITGKPIKFTGTGEKITDIETFHPDRMSSRILGMGD 300

Query: 301 LLTLIERASQEYDEKRSMELAEKMRENTFDFNDFIDQLDQVQNMGPMEDLLKMLPGMANN 360
           LLTLIE+ASQ+YDE++S ELAEKMREN+FDFNDFI+QLDQVQNMG MED+LKM+PGMANN
Sbjct: 301 LLTLIEKASQDYDEQKSAELAEKMRENSFDFNDFIEQLDQVQNMGSMEDILKMIPGMANN 360
```

```
Query: 361 PAMKNFKVDENEIARKRAIVSSMTPEERENPDLLNPSRRRRIAAGSGNTFVDVNKFIKDF 420
           PA+ N +VDE EIARKRAIVSSMTPEERENPDLL PSRRRRIA+GSGNTFV+VNKFIKDF
Sbjct: 361 PALANVEVDEGEIARKRAIVSSMTPEERENPDLLTPSRRRRIASGSGNTFVNVNKFIKDF 420

Query: 421 NQAKQMMQGVMSGDMNKMMKKMGIDPNNLPKDMPGMDGMDMSNLEGMMGQNGMPDLSSL- 479
           NQAK+MMQGVMSGDMNK+MK+MGI+PNN+P      + MD S LEGMMGQ GMPD+S L
Sbjct: 421 NQAKKMMQGVMSGDMNKVMKQMGINPNNMP------NNMDSSALEGMMGQGGMPDMSGLS 474

Query: 480 GGDMDFSQMFGGGLKGKVGAFAAKQSMKRMANKMKKAKKKRK 521
           G +MD SQMFGGGLKGKVG FA KQSMK+MA +MKKAKK++K
Sbjct: 475 GANMDVSQMFGGGLKGKVGEFAMKQSMKKMAKRMKKAKKRKK 516
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2089> which encodes the amino acid sequence <SEQ ID 2090>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22   Transmembrane   39 -  55 (  39 -  55)

----- Final Results -----
               bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 458/522 (87%), Positives = 489/522 (92%), Gaps = 4/522 (0%)

Query: 1   MAFESLTERLQGVFKNIRGKKKLSEKDVQEVTKEIRLALLEADVALPVVKTFIKHVRERA 60
           MAFESLT+RLQ VFK+IRGKKKLSE DVQEVTKEIRLALLEADVALPVVKTFIK VRERA
Sbjct: 3   MAFESLTQRLQDVFKHIRGKKKLSESDVQEVTKEIRLALLEADVALPVVKTFIKRVRERA 62

Query: 61  VGHEIIDTLDPTQQIVKIVNEELTDLLGAETSEIEKSPKIPTIIMMVGLQGAGKTTFAGK 120
           +GHEIIDTLDPTQQI+KIVNEELT +LG+ET+EI+KSPKIPTIIMMVGLQGAGKTTFAGK
Sbjct: 63  IGHEIIDTLDPTQQILKIVNEELTSILGSETAEIDKSPKIPTIIMMVGLQGAGKTTFAGK 122

Query: 121 LANKLIKEDNARPMMIAADIYRPAAIDQLKTLGSQINVPVFDMGTNHSAVEIVTKGLEQA 180
           LANKLIKE+NARP+MIAADIYRPAAIDQLKTLG QINVPVFDMGT+HSAV+IV KGLEQA
Sbjct: 123 LANKLIKEENARPLMIAADIYRPAAIDQLKTLGQQINVPVFDMGTDHSAVDIVRKGLEQA 182

Query: 181 RENRNDYVLIDTAGRLQIDATLMQELHDVKAIAQPNEILLVVDSMIGQEAANVAEEFNRQ 240
           REN NDYVLIDTAGRLQID  LM EL DVKA+AQPNEILLVVDSMIGQEAANVA EFN Q
Sbjct: 183 RENHNDYVLIDTAGRLQIDEKLMGELRDVKALAQPNEILLVVDSMIGQEAANVAYEFNHQ 242

Query: 241 LSISGVVLTKIDGDTRGGAALSVREITGKPIKFTGTGEKITDIETFHPDRMASRILGMGD 300
           LSI+GVVLTKIDGDTRGGAALSVREITGKPIKFTG GEKITDIETFHPDRM+SRILGMGD
Sbjct: 243 LSITGVVLTKIDGDTRGGAALSVREITGKPIKFTGIGEKITDIETFHPDRMSSRILGMGD 302

Query: 301 LLTLIERASQEYDEKRSMELAEKMRENTFDFNDFIDQLDQVQNMGPMEDLLKMLPGMANN 360
           LLTLIE+ASQEYDEK+S+ELAEKMRENTFDFNDFI+QLDQVQNMGPMEDLLKM+PGMA N
Sbjct: 303 LLTLIEKASQEYDEKKSLELAEKMRENTFDFNDFIEQLDQVQNMGPMEDLLKMIPGMAGN 362

Query: 361 PAMKNFKVDENEIARKRAIVSSMTPEERENPDLLNPSRRRRIAAGSGNTFVDVNKFIKDF 420
           PA+ N KVDEN+IARKRAIVSSMTP ERENPDLLNPSRRRRIAAGSGN+FVD NKFIKDF
Sbjct: 363 PALANIKVDENQIARKRAIVSSMTPAERENPDLLNPSRRRRIAAGSGNSFVD-NKFIKDF 421

Query: 421 NQAKQMMQGVMSGDMNKMMKKMGIDPNNLPKDMPGMDGM-DMSNLEGMMGQNGMPDLSSL 479
           NQAK MMQGVMSGDM+KMMK MGI+PNNLPK+MP   GM DMS+LEGMMGQ GMPDLS L
Sbjct: 422 NQAKSMMQGVMSGDMSKMMKDMGINPNNLPKNMPA--GMPDMSSLEGMMGQGGMPDLSGL 479

Query: 480 GGDMDFSQMFGGGLKGKVGAFAAKQSMKRMANKMKKAKKKRK 521
           GGDMD SQ+FG G KGK+G FA KQ+MKR ANK+KKAKKKRK
Sbjct: 480 GGDMDMSQLFGKGFKGKIGQFAMKQAMKRQANKLKKAKKKRK 521
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 679

A DNA sequence (GBSx0721) was identified in *S.agalactiae* <SEQ ID 2091> which encodes the amino acid sequence <SEQ ID 2092>. This protein is predicted to be SatD. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.28   Transmembrane    3 -  19 (   2 -  19)

----- Final Results -----
               bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9811> which encodes amino acid sequence <SEQ ID 9812> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG28336 GB:U88582 SatD [Streptococcus mutans]
 Identities = 106/222 (47%), Positives = 162/222 (72%), Gaps = 2/222 (0%)

Query: 13  MYLALIGDIINSKQILERETFQQSFQQLMTELSDVYGEELISPFTITAGDEFQALLKPSK 72
           +Y+A+IGD+I+SK I  R   Q+  + L+ +++  Y E L S FTIT GDEFQALL P+
Sbjct: 2   IYIAIIGDLISSKAITNRPKSQKQLKNLLNQINKKYKELLKSAFTITTGDEFQALLVPNP 61

Query: 73  KVFQIIDHIQLALKPVNVRFGLGTGNIITSINSNESIGADGPAYWHARSAINHIHDKNDY 132
           ++FQIID I L  KP  +RFG+G+G+I+T IN  +SIG+DGPAYWHAR+AI++IHDKNDY
Sbjct: 62  QIFQIIDEIALGFKPYQIRFGVGSGSILTEINPEQSIGSDGPAYWHARAAIDYIHDKNDY 121

Query: 133 GTVQVAICLDDEDQNLELTLNSLISAGDFIKSKWTTNHFQMLEHLILQDNYQEQFQHQKL 192
           G+  +A+ L+D + + +  +N++++A +FIKSKWT    +++++ L+    Y+E+F H+K+
Sbjct: 122 GSNHLAVDLEDTETSQQ--INAILAACEFIKSKWTVTQYEVIDGLLQAGIYEEKFSHKKM 179

Query: 193 AQLENIEPSALTKRLKASGLKIYLRTRTQAADLLVKSCTQTK 234
           A+  ++ PS+  KRLK+SGLKIYLR +  A  LL+ +  + K
Sbjct: 180 AEKLDLSPSSFNKRLKSSGLKIYLRNKKVATTLLLNAIRKEK 221
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2093> which encodes the amino acid sequence <SEQ ID 2094>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3744(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 94/213 (44%), Positives = 137/213 (64%), Gaps = 3/213 (1%)

Query: 14  YLALIGDIINSKQILERETFQQSFQQLMTELSDVYGEELISPFTITAGDEFQALLKPSKK 73
           Y+ALIGDII SKQ+ +R   Q++    + +L+  +   +IS  ++T GDEFQ L +
Sbjct: 3   YIALIGDIIQSKQLTDRSKVQKTLAAYLDDLNKTFAPYIISKLSLTLGDEFQGLFQVDTP 62

Query: 74  VFQIIDHIQLALKPVNVRFGLGTGNIITSINSNESIGADGPAYWHARSAINHIHDKNDYG 133
           +F +ID I   +  + +RFG+G G+I+T IN + SIGADGPAYWHAR AI +IH KNDYG
Sbjct: 63  IFHLIDLINHHMD-IPIRFGVGVGSILTDINPDISIGADGPAYWHAREAIRYIHQKNDYG 121

Query: 134 TVQVAICLDDEDQNLELTLNSLISAGDFIKSKWTTNHFQMLEHLILQDNYQEQFQHQKLA 193
              +A  L     N +  LNSL++AGD IK+ W  + +++ + L+    Y+E F  Q+L
Sbjct: 122 NTTLA--LRTGHHNQDDVLNSLLAAGDAIKANWRASQWEIFDTLLDLGIYEEYFDQQRLG 179
```

```
Query: 194 QLENIEPSALTKRLKASGLKIYLRTRTQAADLL 226
           +  ++   SAL+KRLK+S +KIYLRTR  A + L
Sbjct: 180 KQLSLSSSALSKRLKSSHVKIYLRTRQSALNCL 212
```

A related GBS gene <SEQ ID 8637> and protein <SEQ ID 8638> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:       4.96
GvH: Signal Score (-7.5): -5.46
     Possible site: 49
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -1.28 threshold:  0.0
   INTEGRAL    Likelihood = -1.28   Transmembrane    3 -  19 (   1 -  19)
   PERIPHERAL  Likelihood =  5.99      74
 modified ALOM score:   0.76

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 8638 (GBS338) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 5; MW 30kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 11; MW 55kDa).

GBS338-GST was purified as shown in Figure 215, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 680

A DNA sequence (GBSx0722) was identified in *S.agalactiae* <SEQ ID 2095> which encodes the amino acid sequence <SEQ ID 2096>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.6082(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 681

A DNA sequence (GBSx0723) was identified in *S.agalactiae* <SEQ ID 2097> which encodes the amino acid sequence <SEQ ID 2098>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have a cleavable N-term signal seq.
```

```
   INTEGRAL    Likelihood = -9.87   Transmembrane  126 - 142 ( 124 - 154)
   INTEGRAL    Likelihood = -8.23   Transmembrane   45 -  61 (  41 -  66)
   INTEGRAL    Likelihood = -5.10   Transmembrane  241 - 257 ( 236 - 257)
   INTEGRAL    Likelihood = -4.04   Transmembrane  199 - 215 ( 198 - 218)
   INTEGRAL    Likelihood = -0.22   Transmembrane   96 - 112 (  96 - 112)

----- Final Results -----
               bacterial membrane --- Certainty=0.4949(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG28337 GB:U88582 SatE [Streptococcus mutans]
 Identities = 54/103 (52%), Positives = 70/103 (67%), Gaps = 2/103 (1%)

Query: 1    MISDFLRDNPILTLLFCAHFLADFQWQSQSLADSKSHSWRGLWRHLLIVFLPLAALMILI 60
            +IS FL  NP+LTLL  AHFLADFQWQSQ +AD KS +W  L RHL+IV LPL  L ++I
Sbjct: 6    VISQFLSGNPVLTLLLIAHFLADFQWQSQKMADLKSSNWTYLIRHLIIVALPLILLSVVI 65

Query: 61   PETTLLNLSIWGSHIVIDSIKKLSYPWVEEGHF--QKAAFIID 101
            P + L+   I+ SH++IDS K L   + ++  F   KA F+ID
Sbjct: 66   PHSFLVLSLIFLSHVLIDSGKLLLNSFYKDRSFIKTKAVFLID 108
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2099> which encodes the amino acid sequence <SEQ ID 2100>. Analysis of this protein sequence reveals the following:

```
     Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.59   Transmembrane  125 - 141 ( 120 - 144)
   INTEGRAL    Likelihood = -6.58   Transmembrane  222 - 238 ( 215 - 238)
   INTEGRAL    Likelihood = -5.04   Transmembrane   47 -  63 (  45 -  77)
   INTEGRAL    Likelihood = -4.62   Transmembrane  179 - 195 ( 178 - 199)
   INTEGRAL    Likelihood = -0.43   Transmembrane   67 -  83 (  67 -  83)

----- Final Results -----
               bacterial membrane --- Certainty=0.4036(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 109/256 (42%), Positives = 146/256 (56%), Gaps = 28/256 (10%)

Query: 2    ISDFLRDNPILTLLFCAHFLADFQWQSQSLADSKSHSWRGLWRHLLIVFLPLAALMILIP 61
            +S +L   P LTL    H L+D+Q QSQ +AD K      L  HL+ V +PL  L ++IP
Sbjct: 5    VSHYLAQTPTLTLFLICHVLSDYQLQSQQVADLKEKHLTYLGYHLIGVSIPLICLTLIIP 64

Query: 62   ETTLLNLSIWGSHIVIDSIKKL---SYPWVEEGHFQKAAFIIDQLAHYTCIIVFYHALPT 118
            +  L++L +  SH +ID +K     S  W  E       F++DQ  H          L
Sbjct: 65   QAWLMSLLVMISHALIDWLKPKMANSLKWKREW-----IFLLDQCLHIAISSFAGLRLAG 119

Query: 119 YLPPNHWLLPIKHFIVIALVFIIITKPINIVFKIFFNKFQAKELSSLLTQEKTKIMKEKS 178
               PN WL PI   ++  L  ++ITKP NIVFK+FF K+Q  +   +
Sbjct: 120 VTLPN-WL-PIS-ILMTVLFILLITKPTNIVFKLFFIKYQPDQGEKM------------- 163

Query: 179 EDHEETIEGAGAMIGNLERLIMAILLISGQYAAIGLVFTAKSIARYDKISKSQVFAEYYL 238
                +TI GAGA IG LER+++ + +I GQ+A+IGLVFTAKSIARY+KIS+S  FAEYYL
Sbjct: 164 ----DTIIGAGATIGILERIVIGVCMIMGQFASIGLVFTAKSIARYNKISESPAFAEYYL 219

Query: 239 IGSLFSIISVLITHWL 254
            IGSLFSI+SV I  W+
Sbjct: 220 IGSLFSILSVFIAAWI 235
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 682

A DNA sequence (GBSx0724) was identified in *S.agalactiae* <SEQ ID 2101> which encodes the amino acid sequence <SEQ ID 2102>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD17886 GB:AF100456 hyaluronate-associated protein precursor
           [Streptococcus equi]
 Identities = 358/521 (68%), Positives = 426/521 (81%), Gaps = 2/521 (0%)

Query: 1   MSSFNRKKLKFLGISLATLTATTVTLVACGNESKNSGDNKV-INWYIPTEISTLDISKNT 59
           M+    K  K LG++  TL A+   L+ACGN+   S D K  INWY PTEI TLDISKNT
Sbjct: 1   MTVLGTKACKRLGLAAVTL-ASVAALMACGNKQSASTDKKSEINWYTPTEIITLDISKNT 59

Query: 60  DAYSNLAIGNSGSNLLRIDKEGKPKPDLAKKVSVSSDGLTYTATLRDNLKWSDGSKLSAE 119
           D YS LAIGNSGSNLLR D +GK +PDLA+KV VS DGLTYTATLRD LKWSDGS L+AE
Sbjct: 60  DTYSALAIGNSGSNLLRADAKGKLQPDLAEKVDVSEDGLTYTATLRDGLKWSDGSDLTAE 119

Query: 120 DFVYTWRRIVDPKTASEYAYLATESHLLNADKINSGDIKDLNKLGVTAKGNQVTFKLTSP 179
           DFVY+W+R+VDPKTASEYAYLATESHL NA+ INSG   DL+ LGV A GN+V F LT P
Sbjct: 120 DFVYSWQRMVDPKTASEYAYLATESHLKNAEDINSGKNPDLDSLGVKADGNKVIFTLTEP 179

Query: 180 CPQFKYYLAFSNFMPQKQSYVEKVGKDYGTTSKNQIYSGPYLVKDWNGSNGKFKLVKNKY 239
            PQFK  L+FSNF+PQK+S+V+  GKDYGTTS+ QIYSGPY+VKDWNG++G FKLVKNK
Sbjct: 180 APQFKSLLSFSNFVPQKESFVKDAGKDYGTTSEKQIYSGPYIVKDWNGTSGTFKLVKNKN 239

Query: 240 YWDSKHVKTNSVIVQTIKKPDTAVQMYKQGQIDFAEISGTSAIYQANKNNKDVVDASDAR 299
           YWD+K+VKT +V VQT+KKPDTAVQMYKQG++DFA ISGTSAIY ANK +KDVV   +A
Sbjct: 240 YWDAKNVKTETVNVQTVKKPDTAVQMYKQGKLDFANISGTSAIYNANKKHKDVVPVLEAT 299

Query: 300 TTYIIYNQTGSVKALTNQKIRQALNLATDRKGVVKAAVDTGSTPAESLVPKKLAKLPNGE 359
           T YI+YNQTG+++ L + KIRQALNLATDRKG+V AAVDTGS PA +LVP  LAKL +G
Sbjct: 300 TAYIVYNQTGAIEGLNSLKIRQALNLATDRKGIVSAAVDTGSKPATALVPTGLAKLSDGT 359

Query: 360 DLSKYTAPGYTYNTSKAQKLFKEGLAEVGQSSLKLTITADSDSPAAKNAVDYVKSTWESA 419
           DL+++ APGY Y+  +A KLFKEGLAE+G+ +L +TITAD+D+PAAK+AVDY+K TWE+A
Sbjct: 360 DLTEHVAPGYKYDDKEAAKLFKEGLAELGKDALTITITADADAPAAKSAVDYIKETWETA 419

Query: 420 LPGLTVEEKFVTFKQRLEDAKNENFDVVLFSWGGDYPEGSTFYGLFTTNSAYNYGKFSSK 479
           LPGLTVEEKFV FKQRLED KN+NF+V +   WGGDYP+GSTFYGLF + SAYNYGKF++
Sbjct: 420 LPGLTVEEKFVPFKQRLEDTKNQNFEVAVVLWGGDYPKGSTFYGLFKSGSAYNYGKFTNA 479

Query: 480 EYDNAYQKAITTDALKPGDAANDYKTAEKALFDQSYYNPVY 520
           +YD AY KA+TTDAL    AA+DYK AEKAL+D + YNP+Y
Sbjct: 480 DYDAAYNKALTTDALNTDAAADDYKAAEKALYDNALYNPLY 520
```

There is also homology to SEQ ID 318. An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 138/524 (26%), Positives = 222/524 (42%), Gaps = 73/524 (13%)

Query: 7   KKLKFLG-ISLATLTATTVTLVACGNESKNSGDN--KVINWYIPTEISTLDISKNTDAYS 63
           KK K+L  +S+A L+ +   L ACGN++ + G    K   +    +  +LD        +
Sbjct: 5   KKSKWLAAVSVAILSVSA--LAACGNKNASGGSEATKTYKYVFVNDPKSLDYILTNGGGT 62
```

```
Query: 64  NLAIGNSGSNLLRIDKEGKPKPDLAKKVSVSSDGLTYTATLRDNLKW--SDGSK---LSA 118
              I      LL  D+ G   P LAK   VS DGLTYT TLRD + W  +DG +   ++A
Sbjct: 63  TDVITQMVDGLLENDEYGNLVPSLAKDWKVSKDGLTYTYTLRDGVSWYTADGEEYAPVTA 122

Query: 119 EDFVYTWRRIVDPKTASEYAYLATESHLLNADKINSGDIKDLNKLGVTAKGNQ-VTFKLT 177
           EDFV   +  VD K+ + Y     E  + N     +G++ D  ++GV A  ++ V + L
Sbjct: 123 EDFVTGLKHAVDDKSDALY---VVEDSIKNLKAYQNGEV-DFKEVGVKALDDKTVQYTLN 178

Query: 178 SPCPQFKYYLAFSNFMPQKQSYVEKVGKDYGTTSKNQI-YSGPYLVKDWNGSNGKFKLVK 236
            P   +     +S   P    +++  GKD+GTT  + I  +G Y +  + S    + K
Sbjct: 179 KPESYWNSKTTYSVLFPVNAKFLKSKGKDFGTTDPSSILVNGAYFLSAFT-SKSSMEFHK 237

Query: 237 NKYYWDSKHVKTNSV--IVQTIKKPDTAVQMYKQGQIDFAEISGTSAIYQ-ANKNNKDVV 293
           N+ YWD+K+V   SV         P +  + + +G+   A +      Y+ A KN  D +
Sbjct: 238 NENYWDAKNVGIESVKLTYSDGSDPGSFYKNFDKGEFSVARLYPNDPTYKSAKKNYADNI 297

Query: 294 D----ASDARTTYIIYN---------------QTGSVKALTNQKIRQALNLATDRKG--- 331
                  D R  ++ +N               Q    KAL N+  RQA+  A DR
Sbjct: 298 TYGMLTGDIR--HLTWNLNRTSFKNTKKDPAQQDAGKKALNNKDFRQAIQFAFDRASFQA 355

Query: 332 ----------------VVKAAVDTGSTPAESLVPKKLAKL-PNGEDLSKYTAPGYTYNTS 374
                           V    V  G +   S V K++AKL    +D++   A    YN
Sbjct: 356 QTAGQDAKTKALRNMLVPPTFVTIGESDFGSEVEKEMAKLGDEWKDVNLADAQDGFYNPE 415

Query: 375 KAQKLF---KEGLAEVGQS-SLKLTITADSDSPAAKNAVDYVKSTWESALPGLTV----- 425
           KA+  F   KE L   G +  ++L    D  + A        K + E++L    V
Sbjct: 416 KAKAEFAKAKEALTAEGVTFPVQLDYPVDQANAATVQEAQSFKQSVEASLGKENVIVNVL 475

Query: 426 EEKFVTFKQR---LEDAKNENFDVVLFSWGGDYPEGSTFYGLFT 466
           E +  T + +    E  + +++D++   WG DY +  T+  + +
Sbjct: 476 ETETSTHEAQGFYAETPEQQDYDIISSWWGPDYQDPRTYLDIMS 519
```

SEQ ID 2102 (GBS323) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 4; MW 61.3kDa).

The GBS323-His fusion product was purified (Figure 209, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 306), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 683

A DNA sequence (GBSx0725) was identified in *S.agalactiae* <SEQ ID 2103> which encodes the amino acid sequence <SEQ ID 2104>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.54   Transmembrane  199 - 215 ( 198 - 215)

----- Final Results -----
               bacterial membrane --- Certainty=0.1617(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC17173 GB:AF065141 unknown [Streptococcus mutans]
 Identities = 304/356 (85%), Positives = 334/356 (93%)

Query: 1   MKRELLLEKIDELKEIMPWYVLEYYQSKLSVPYSFTTLYEYLKEYRRFLEWLLDSGVANC 60
           M+RELLLEKIDELKE+MPWYVLEYYQSKL+VPYSFTTLYEYLKEYRRF EWL+DSGV+N
Sbjct: 1   MRRELLLEKIDELKELMPWYVLEYYQSKLTVPYSFTTLYEYLKEYRRFFEWLIDSGVSNA 60
```

```
Query: 61  HHIAEIELSVLENLTKKDMEAFILYLRERPLLNANTRQNGVSQTTINRTLSALSSLFKYL 120
           + +A+I L  LE+L+KKDME+FILYLRER LLN   ++ GVSQTTINRTLSALSSL+KYL
Sbjct: 61  NKLADIPLETLEHLSKKDMESFILYLRERTLLNTKNKRQGVSQTTINRTLSALSSLYKYL 120

Query: 121 TEEVENADGEPYFYRNVMKKVSTKKKKETLASRAENIKQKLFLGNETIEFLEYIDCEYQN 180
           TEEVENADGEPYFYRNVMKKVSTKKKKETLA+RAENIKQKLFLGNET+EFLEY+DCEY+
Sbjct: 121 TEEVENADGEPYFYRNVMKKVSTKKKKETLAARAENIKQKLFLGNETMEFLEYVDCEYEQ 180

Query: 181 KLSKRALAFFNKNKERDLAIIALLLASGVRLSEAVNLDLKDINLNVMVIDVTRKGGKRDS 240
           KLSKRAL+ F KNKERDLAIIALLLASGVRLSEAVNLDLKD+NLN+M+I+VTRKGGK DS
Sbjct: 181 KLSKRALSSFRKNKERDLAIIALLLASGVRLSEAVNLDLKDVNLNMMIIEVTRKGGKHDS 240

Query: 241 VNVASFAKPYLANYLDIRKNRYKAENQDIALFLSEYRGVPNRIDASSVEKMVAKYSQDFK 300
           VNVA FAKPYL NY+ IR+ RYKA+  D+A FLSEYRGVPNR+DASS+EKMVAKYSQDFK
Sbjct: 241 VNVAGFAKPYLENYITIRRGRYKAKKTDLAFFLSEYRGVPNRMDASSIEKMVAKYSQDFK 300

Query: 301 VRVTPHKLRHTLATRLYDATKSQVLVSHQLGHASTQVTDLYTHIVNDEQKNALDKL 356
           +RVTPHKLRHTLATRLYDATKSQVLVSHQLGHASTQVTDLYTHIVNDEQKNALDKL
Sbjct: 301 IRVTPHKLRHTLATRLYDATKSQVLVSHQLGHASTQVTDLYTHIVNDEQKNALDKL 356
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2105> which encodes the amino acid sequence <SEQ ID 2106>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.54   Transmembrane  211 - 227 ( 210 - 227)

----- Final Results -----
               bacterial membrane --- Certainty=0.1617(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9139> which encodes the amino acid sequence <SEQ ID 9140>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 60
>>> Seems to have no N-terminal signal sequence
     INTEGRAL    Likelihood = -1.54   Transmembrane  199 - 215 ( 198 - 215)

----- Final Results -----
               bacterial membrane --- Certainty= 0.162(Affirmative) < succ>
                bacterial outside --- Certainty= 0.000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 283/356 (79%), Positives = 321/356 (89%)

Query: 1   MKRELLLEKIDELKEIMPWYVLEYYQSKLSVPYSFTTLYEYLKEYRRFLEWLLDSGVANC 60
           M+RELLLEKI+  K IMPWYVL+YYQSKL+VPYSFTTLYEYLKEY+RF +WL+D+ +
Sbjct: 13  MRRELLLEKIETYKAIMPWYVLDYYQSKLAVPYSFTTLYEYLKEYKRFFDWLMDADLTQA 72

Query: 61  HHIAEIELSVLENLTKKDMEAFILYLRERPLLNANTRQNGVSQTTINRTLSALSSLFKYL 120
             IA+I+LS LE+LTKKD+EAF+LYLRERP LN  + + G+SQTTINRTLSALSSL+KYL
Sbjct: 73  PKIADIDLSTLEHLTKKDLEAFVLYLRERPSLNTYSTKEGLSQTTINRTLSALSSLYKYL 132

Query: 121 TEEVENADGEPYFYRNVMKKVSTKKKKETLASRAENIKQKLFLGNETIEFLEYIDCEYQN 180
           TEEVEN  GEPYFYRNVMKKVSTKKKKETLASRAENIKQKLFLG+ET+ FL+Y+D EY+
Sbjct: 133 TEEVENDQGEPYFYRNVMKKVSTKKKKETLASRAENIKQKLFLGDETLAFLDYVDKEYEQ 192

Query: 181 KLSKRALAFFNKNKERDLAIIALLLASGVRLSEAVNLDLKDINLNVMVIDVTRKGGKRDS 240
           KLS RA + F KNKERDLAIIALLLASGVRLSEAVNLDLKD+NLN+M+I+V RKGGKRDS
Sbjct: 193 KLSNRAKSSFRKNKERDLAIIALLLASGVRLSEAVNLDLKDVNLNMMIIEVIRKGGKRDS 252

Query: 241 VNVASFAKPYLANYLDIRKNRYKAENQDIALFLSEYRGVPNRIDASSVEKMVAKYSQDFK 300
```

```
           VNVA FAK YL +YL +R+ RYKAE QD+A FL+EYRGVPNR+DASS+EKMV KYS+DFK
Sbjct: 253 VNVAGFAKGYLESYLAVRQRRYKAEKQDLAFFLTEYRGVPNRMDASSIEKMVGKYSEDFK 312

Query: 301 VRVTPHKLRHTLATRLYDATKSQVLVSHQLGHASTQVTDLYTHIVNDEQKNALDKL 356
           +RVTPHKLRHTLATRLYDATKSQVLVSHQLGH+STQVTDLYTHIVNDEQKNALD L
Sbjct: 313 IRVTPHKLRHTLATRLYDATKSQVLVSHQLGHSSTQVTDLYTHIVNDEQKNALDNL 368
```

SEQ ID 2104 (GBS420) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 5; MW 68kDa).

GBS420-GST was purified as shown in Figure 219, lane 9-10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 684

A DNA sequence (GBSx0726) was identified in *S.agalactiae* <SEQ ID 2107> which encodes the amino acid sequence <SEQ ID 2108>. This protein is predicted to be a sensor-like histidine kinase in idh 3'region. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.75   Transmembrane    10 -  26 (   8 -  34)
   INTEGRAL    Likelihood = -3.93   Transmembrane    37 -  53 (  35 -  54)

----- Final Results -----
              bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB16001 GB:Z99124 similar to two-component sensor histidine
           kinase [YxdJ] [Bacillus subtilis]
 Identities = 96/320 (30%), Positives = 172/320 (53%), Gaps = 16/320 (5%)

Query: 2   IRQFLREHLIWYILYIM--MFVLFFISFYLYHLPMPYLFNSLGLNVIVLLGISIWQYSRY 59
           ++ FLR H +  +L+++  +FV F+  F   H    +LF  LG+ +++L G   +++ +
Sbjct: 1   MKLFLRSHAVLILLFLLQGLFVFFYYWFAGLH-SFSHLFYILGVQLLILAGYLAYRWYKD 59

Query: 60  RKKMLHLKYFNSSQDPSFELQPSDYAYFNIITQLEA--REAQKVSETIEQTNHVALMIKM 117
           R     L       D  + L  S +       Q+E    + QK+ ET  + +     +
Sbjct: 60  RGVYHWLSSGQEGTDIPY-LGSSVFCSELYEKQMELIRLQHQKLHETEAKLDARVTYMNQ 118

Query: 118 WSHQMKVPLAAISLMAQTNHLDP--KEVEQQLLKLQHYLETLLAFLKFRQYRDDFRFEAV 175
           W HQ+K PL+ I+L+ Q    +P  +++++++ +++  LETLL   +   +  DF+ EAV
Sbjct: 119 WVHQVKTPLSVINLIIQEED-EPVFEQIKKEVRQIEFGLETLLYSSRLDLFERDFKIEAV 177

Query: 176 SLREVVVEIIKSYKVICLSKSL--SIIIEGDNIWKTDKKWLTFALSQVLDNAIKYSNPES 233
           SL E++  +I+SYK   +   +   + +  D+   TD KWL FA+ QV+ NA+KYS  +S
Sbjct: 178 SLSELLQSVIQSYKRFFIQYRVYPKMNVCDDHQIYTDAKWLKFAIGQVVTNAVKYSAGKS 237

Query: 234 -----KIIISIGEESIRIQDYGIGILEEDIPRLFEDGFTGYNGHEHQKATGMGLYMTKEV 288
                 +        + ++DYG+GI  +DI R+F+  +TG NG   Q++TG+GL++ KE+
Sbjct: 238 DRLELNVFCDEDRTVLEVKDYGVGIPSQDIKRVFDPYYTGENGRRFQESTGIGLHLVKEI 297

Query: 289 LSSLNLSISVDSKINYGTAV 308
              LN ++ + S    GT+V
Sbjct: 298 TDKLNHTVDISSSPGEGTSV 317
```

SEQ ID 2108 (GBS421) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 6; MW 63kDa).

GBS421-GST was purified as shown in Figure 219, lane 11.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 685

A DNA sequence (GBSx0727) was identified in *S.agalactiae* <SEQ ID 2111> which encodes the amino acid sequence <SEQ ID 2112>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1310(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD10258 GB:AF036964 putative response regulator [Lactobacillus
          sakei]
 Identities = 94/222 (42%), Positives = 140/222 (62%), Gaps = 8/222 (3%)

Query: 7   KIYIVEDDMTIVSLLKDHLSASYHVSSV--SNFRDVKQEIIAFQPDLILMDITLPYFNGF 64
            +I IVEDD TI +L+ ++L   + + ++   +F  +     +  +P L+L+DI LP ++GF
Sbjct: 3   EIMIVEDDPTIANLIAENLE-KWQLKAIIPDDFDTIFDRFLTDKPHLVLLDINLPVYDGF 61

Query: 65  YWTAELRKFLTIPIIFISSSNDEMDMVMALNMGGDDFISKPFSLAVLDAKLTAILRRSQQ 124
            YW  ++R+   +PIIFISS +  MDMVM++NMGGDDF++KPFS+ VL AK+ A+LRR+
Sbjct: 62  YWCRKIREVSKVPIIFISSRSTNMDMVMSMNMGGDDFVNKPFSMEVLIAKINALLRRTYN 121

Query: 125 FIQQE---LTFGGFTLT-REGLLSSQDKEVILSPTENKILSILLMHPKQVVSKESLLEKL 180
            ++ Q    +   G  +  + G     D  V LS  E K+L  L+    Q+VS+E LL  L
Sbjct: 122 YVDQNTDVIEHNGLLINLQSGGAQVGDTVVDLSKNEYKLLQFLMRQHGQIVSREKLLRAL 181

Query: 181 WENDSFIDQNTLNVNMTRLRKKIVPIGF-DYIHTVRGVGYLL 221
            W+++ F+D NTL VN+ RLRKKI   G  DYI T  G GY++
Sbjct: 182 WDDERFVDDNTLTVNINRLRKKIEQAGLEDYIQTKIGQGYII 223
```

There is also homology to SEQ ID 1182.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 686

A DNA sequence (GBSx0728) was identified in *S.agalactiae* <SEQ ID 2113> which encodes the amino acid sequence <SEQ ID 2114>. This protein is predicted to be permease OrfY. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.62   Transmembrane   55 -  71 (  49 -  75)
   INTEGRAL    Likelihood =-10.30   Transmembrane  197 - 213 ( 192 - 218)
   INTEGRAL    Likelihood = -9.13   Transmembrane  152 - 168 ( 141 - 172)
   INTEGRAL    Likelihood = -8.70   Transmembrane  624 - 640 ( 619 - 645)
   INTEGRAL    Likelihood = -8.44   Transmembrane  222 - 238 ( 219 - 250)
   INTEGRAL    Likelihood = -7.75   Transmembrane  283 - 299 ( 280 - 307)
   INTEGRAL    Likelihood = -7.70   Transmembrane  533 - 549 ( 526 - 552)
   INTEGRAL    Likelihood = -6.95   Transmembrane  108 - 124 (  99 - 140)
   INTEGRAL    Likelihood = -4.88   Transmembrane  585 - 601 ( 581 - 610)
   INTEGRAL    Likelihood = -3.82   Transmembrane   25 -  41 (  21 -  47)
```

```
    INTEGRAL    Likelihood = -0.48   Transmembrane  602 - 618 ( 602 - 618)

----- Final Results -----
                bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9809> which encodes amino acid sequence <SEQ ID 9810> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF99695 GB:AF267498 permease OrfY [Streptococcus mutans]
 Identities = 154/665 (23%), Positives = 299/665 (44%), Gaps = 40/665 (6%)

Query: 4   MFYLKIAWHNLKHSIDQYIPFLLASLLLYSLTCSTLLILMSAVGRDMGTAAT---VLFLG 60
            MF  KI++HNL  +    +P+     +        +  L +    ++ TA     +L  G
Sbjct: 1   MFLPKISFHNLIVNKSLTLPYFAIMTIFSGFNYVLINFLTNPSFYNIPTARILIDILIFG 60

Query: 61  VIVLSIFAVVMEHYSYNILMKQRSSEFGLYNILGMNKRQVARVASLELFIIYIFLISIGS 120
             I++S+  ++   Y+   +  +R+S  G++ +LGM K+Q+ ++  LE   ++      G
Sbjct: 61  FILISLLMLLYGRYANRFISDERNSNMGIFLMLGMGKKQLLKIIYLEKLYLFTGTFFGGL 120

Query: 121 LFSAFFAKFIYLIFVNIINYHALNLSLSLWPFIICIVIFTGIFLTLEVPVIRHVHLSSPL 180
            +F   ++K  +L   N+I    +    SL      +++   I+  + +   R +   S
Sbjct: 121 IFGFVYSKIFFLFIRNLIVIGDVREQYSLTAISWLLILTFFIYFIIYLSEYRLLKRQSIT 180

Query: 181 SLFRKKQQGEKEPKGNLILAILALVAIAIAYTMALTSGKAPALAVIY-RFFFAVLLVIAG 239
             +F  K + +   K ++ + +  L A+ + Y  ALTS   P +   + RF +A  LV  G
Sbjct: 181 VIFNSKAKRDNPRKTSVFVGLFGLFALLMGYHFALTS---PNVTTSFSRFIYAACLVTLG 237

Query: 240 TYLFYISFMTWYLKRLRQNKHYYYKSEHFVSTSQMIFRMKQNAVGLASITLLAVMALVTI 299
             +  + S +   L  +++ +  YY    FV + +   R++ NA+ LA+I + +   LV++
Sbjct: 238 IFCTFSSGVIMLLTVIKKRRAIYYNQRRFVVIASLFHRIRSNALSLATICIFSTATLVSL 297

Query: 300 ATTVSLYSNTQNVVTGLFPKSVSLSIDNSKGDAKNIFEEKILKKLGKSSKEAITYNQTMI 359
            +   SLY    N+V    P+ V++    S  D      E  L  +   +   +T  Q
Sbjct: 298 SVLASLYLAKDNMVRLSSPRDVTVL---STTDI-----EPNLMDIATKNHVTLTNRQ--- 346

Query: 360 SMPVSQSSELNITSKNVKHVDITKTGFMY------LITQNDFRRLGHQLPKLKDNQVAYF 413
            ++ VSQS   NI      H+ +   G M       +I+ + F    +   +LK++++  +
Sbjct: 347 NLKVSQSVYGNIKGS---HLSVDPNGGMANDYQITVISLDSFNASNNTHYRLKNHEILTY 403

Query: 414 VQKGDSRLKKINLLGNKFDVVKNLKEA-YVPETTNTYNPGLIIFANNKQI-DNIRKAYLP 471
            V  G +        G K   VK +K   ++     +  P   I  +N++I   I K  L
Sbjct: 404 VSNGAAAPSSYTTNGVKLTNVKQIKRINFIFSPLRSMQPNFFIITDNREIIQTILKEELT 463

Query: 472 YTKNINTFPKTFKAYLDLNSQEINSISKNDIIEVDG--KYVGNISTKQSFLKEGYQMFGG 529
            +         T   Y  +  +++N     D +E     ++  N+ + +        +FGG
Sbjct: 464 WG--------TMAGY-HVKGKKMNQKDFYDELETTNFRQFSANVVSIRQVKSMFNALFGG 514

Query: 530 LLFTGFLLGISFLLGIALIVYYKQYSEGHEDKRSYRILQEVGMSKKLVKRTINSQIMIFF 589
            LLF G + G  F +  A+ +YY+Q SEG  D+  Y+ + ++GM+ K ++ +I  QI   F
Sbjct: 515 LLFVGIIFGTIFAILTAITIYYQQLSEGIRDRDDYKAMIKLGMTNKTIQDSIKVQINFVF 574

Query: 590 FQPLVVAVIHFGVAIPMLKQMLLVFGVLNSTIVYVVSGLTVLAISIIYFIIYRITSRTYY 649
              P+  A+++   A+P+L +++  FG  ++ +     G  ++     Y+ I   TS+ YY
Sbjct: 575 ILPIAFALLNLIFALPILYKIMTTFGFNDAGLFLRAVGTCLIVYLFFYWFICHCTSKLYY 634

Query: 650 HIIER 654
             +I +
Sbjct: 635 RLISK 639
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2115> which encodes the amino acid sequence <SEQ ID 2116>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-13.59   Transmembrane  602 - 618 ( 592 - 630)
   INTEGRAL    Likelihood =-12.26   Transmembrane   59 -  75 (  50 -  81)
   INTEGRAL    Likelihood =-12.21   Transmembrane  235 - 251 ( 224 - 262)
   INTEGRAL    Likelihood = -9.82   Transmembrane  159 - 175 ( 146 - 177)
   INTEGRAL    Likelihood = -9.02   Transmembrane  201 - 217 ( 198 - 223)
   INTEGRAL    Likelihood = -8.97   Transmembrane  510 - 526 ( 507 - 540)
   INTEGRAL    Likelihood = -6.42   Transmembrane  569 - 585 ( 564 - 589)
   INTEGRAL    Likelihood = -5.95   Transmembrane  109 - 125 ( 102 - 138)
   INTEGRAL    Likelihood = -4.09   Transmembrane  294 - 310 ( 290 - 315)
   INTEGRAL    Likelihood = -1.86   Transmembrane  126 - 142 ( 126 - 142)

----- Final Results -----
              bacterial membrane --- Certainty=0.6434(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB03337 GB:AB035452 ABC transporter [Staphylococcus aureus]
 Identities = 141/657 (21%), Positives = 289/657 (43%), Gaps = 66/657 (10%)

Query: 5   ITKSNIKKNFSLYRIYFLATIGLLSIFIAFLNFISDKII--TEKIGDSGQALVIANGSL- 61
           I   N+++N   Y +Y      L S+F + + + S   +  T+ + +     +I  G+L
Sbjct: 6   IVFKNLRQNLKHYAMY------LFSLFFSIVLYFSFTTLQFTKGVNNDDSMAIIKKGALV 59

Query: 62  --IFLIVFLVVFLIYFNNFFVKKRSQELGVLAILGFSKRELTKLLTLENLVILVLSYLVS 119
             IFL + +V+FL+Y N+ FVK+R++E  +  ++G +++ + K+L LE +++ +++ +V
Sbjct: 60  GSIFLFIIIVIFLMYANHLFVKRRTREFALFQLIGLTRQNILKMLALEQMIVFLITGVVG 119

Query: 120 LLLGPTLYFLAVLAITHLLNLTMEVQWFITVNEIIESLGILVVVFLINVITNGLIISKQS 179
           +L G     L +  ++ L++L++ +        ++ ++ +L++ +++ +  + L + ++S
Sbjct: 120 VLCGIAGAQLLLSIVSKLMSLSINLSIHFEPMALVLTIFMLIIAYVLILFQSALFLKRRS 179

Query: 180 LIEFVNFSRKAE----KKIKIRKVRAIIAITALLLSYILCLATVFSSTRNMLLSIGMVPV 235
           ++  +  S K +    K      +  ++ I  + L Y   +AT    T   L      P
Sbjct: 180 ILSMMKDSIKTDATTAKVTTAEVISGVLGIAMIALGYY--MATEMFGTFKALTMAMTSP- 236

Query: 236 SLLIIVLVVLGTVFTIRYGLAFVVSLLKENKKRLYRPLSNIIYPKFNYRIATKNKLLTVL 295
             +I+ L V+G     R  ++ +   LK++K         +      YR+      LT++
Sbjct: 237 -FIILFLTVVGAYLFFRSSVSLIFKTLKKSKNGRVSITDVVFTSSIMYRMKKNAMSLTII 295

Query: 296 GGLLTVTVSVAGMMVMLYAYSLNGIERLTPSAIEYNVESENGQVNVTTILENDQVSL--- 352
             +  VTV+V     +  + +   +  + P+  E+NV +        T L   Q++
Sbjct: 296 AIISAVTVTVLCFAALSKSNTDQTLTSMAPN--EFNVVATQDAKQFETKLSQQQITFSKN 353

Query: 353 ----VDVGLLRLNTIPEVTITDSGQTIPYFDIINYSDYKELMKAQGRTNSIEGSKSLPLL 408
               + V  ++   I     +DSG+T       N        K  G    I  +KSLP +
Sbjct: 354 AYETITVDNVKDQVITLENGSDSGRTNSILSANN--------KVTGNNAIITNTKSLPNI 405

Query: 409 INYYPTEISLGKTFNLGNAYDVT--VKQVSTNNVFSFSTSVTTLV--VSDKLYAKLSSRF 464
           IN     I L K   +    + T  V Q     V+  + S  + V  VS + Y +L +
Sbjct: 406 IN-----IHLNKDLVVKGTKNETFRVTQEDKGRVYPLNLSFNSPVVEVSPEKYQQLKT-- 458

Query: 465 PEKEMTIRTFNGTSIR------SSEAFYNQFSMVPDVISSYSKEHTVKTANIATYIFIT- 517
              +  + TF G  I+       ++A   QF    D + +Y +      A     IF+T
Sbjct: 459 ---QNNVHTFYGYDIKQTSQKEKAQAIAKQFG---DKVITYDEMKKEVDATNGILIFVTS 512

Query: 518 FLSILFIICTGSILYFTSLIEIMENKEEYGYLSKLGYSKKMIHRILRYETGILFLIPVFI 577
           FL + F++  G I+Y   +  E  +    +  L ++G++   + + L  +    F +P+ I
Sbjct: 513 FLGLAFLVAAGCIIYIKQMDETEDELSNFRILKRIGFTHTDMLKGLLLKITFNFGLPLLI 572

Query: 578 GIVNGGMLLIYYKYLFMDTLVAGNIIMLSLLLCLLFFLIIYGTFYVLTLRLVTSIIK 634
            I++     I +  L       GNI  + +++ ++ + +IY TF ++       +IK
Sbjct: 573 AILHAVFAAIAFMKLM------GNISFMPVIVVIVVYTLIYITFALIAFVHSNKLIK 623
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 145/678 (21%), Positives = 277/678 (40%), Gaps = 89/678 (13%)

Query: 13  NLKHSIDQYIPFLLASLLLYSLTCSTL-----LILMSAVGRDMGTAATVLFLGVIVLSIF 67
            N+K +   Y  + LA++ L S+  + L      I+   +G D G A  +    +I L +F
Sbjct: 9   NIKKNFSLYRIYFLATIGLLSIFIAFLNFISDKIITEKIG-DSGQALVIANGSLIFLIVF 67

Query: 68  AVVMEHYSYNILMKQRSSEFGLYNILGMNKRQVARVASLELFIIYIFLISIGSLFSAFFA 127
             VV   Y  N  +K+RS E G+  ILG +KR++ ++ +LE  +I +    +  L S
Sbjct: 68  LVVFLIYFNNFFVKKRSQELGVLAILGFSKRELTKLLTLENLVILV----LSYLVSLLLG 123

Query: 128 KFIYLIFVNIINYHALNLSLSLWPFIICIVIFTGIFLTLEVPVIRHV------HLSSPLS 181
              +Y + V  I  H LNL++ +  FI    I   + + + V +I  +         S +
Sbjct: 124 PTLYFLAVLAIT-HLLNLTMEVQWFITVNEIIESLGILVVVFLINVITNGLIISKQSLIE 182

Query: 182 LFRKKQQGEKEPKGNLILAILALVAIAIAYTMAL------TSGKAPALAVIYRFFFAVLL 235
                 ++ EK+ K   + AI+A+ A+ ++Y + L      T     ++ ++      ++L
Sbjct: 183 FVNFSRKAEKKIKIRKVRAIIAITALLLSYILCLATVFSSTRNMLLSIGMVPVSLLIIVL 242

Query: 236 VIAGTYLFYISFMTWYLKRLRQNKHYYYKSEHFVSTSQMIFRMKQNAVGLASITLLAVMA 295
            V+ GT       + + +  L++NK   Y+    +   +  +R+   A     +T+L  +
Sbjct: 243 VVLGTVFTIRYGLAFVVSLLKENKKRLYRPLSNIIYPKFNYRI---ATKNKLLTVLGGLL 299

Query: 296 LVTIATT---VSLYSNTQNVVTGLFPKSVSLSIDNSKGDAKNIFEEKILKKLGKSSKEAI 352
             VT++     V LY+ + N +  L P ++  ++++  G                 +   I
Sbjct: 300 TVTVSVAGMMVMLYAYSLNGIERLTPSAIEYNVESENGQV---------------NVTTI 344

Query: 353 TYNQTMISMPVSQSSELNITSKNVKHVDITKTG----FMYLITQNDFRRL------GHQL 402
              N  +  + V       +    +  V IT +G    +  +I  +D++ L       + +
Sbjct: 345 LENDQVSLVDVGL-----LRLNTIPEVTITDSGQTIPYFDIINYSDYKELMKAQGRTNSI 399

Query: 403 PKLKDNQVAYFVQKGDSRLKKINLLGNKFDVVKNLKEAYVPETTNTYNPGLIIFANNKQI 462
               K   +       +  L K   LGN +DV   +K+       +       +  ++K
Sbjct: 400 EGSKSLPLLINYYPTEISLGKTFNLGNAYDVT--VKQVSTNNVFSFSTSVTTLVVSDKLY 457

Query: 463 DNIRKAYLPYTKNINTFPKT-------FKAYLDLNSQEINSISKNDIIEVDGKYVGNIST 515
              +   +      I TF  T       F     +    I+S SK   ++       NI+T
Sbjct: 458 AKLSSRFPEKEMTIRTFNGTSIRSSEAFYNQFSMVPDVISSYSKEHTVKT-----ANIAT 512

Query: 516 KQSFLKEGYQMFGGLLFTGFLLGISFLLGIALIVYYKQYSEGHEDKRSYRILQEVGMSKK 575
                           +F  FL  I F++    I+Y+    E  E+K  Y  L ++G SKK
Sbjct: 513 --------------YIFITFL-SILFIICTGSILYFTSLIEIMENKEEYGYLSKLGYSKK 557

Query: 576 LVKRTINSQIMIFFFQPLVVAVIHFGVAIPMLKQMLLVFGVLNSTIVYVVSGLTVLAISI 635
            ++ R +  +  I F  P+ + +++ G+ +   K  L +  ++   I+ +   L +L   I
Sbjct: 558 MIHRILRYETGILFLIPVFIGIVNGGMLLIYYK-YLFMDTLVAGNIIMLSLLLCLLFFLI 616

Query: 636 IYFIIYRITSRTYYHIIE 653
            IY   Y +T R    II+
Sbjct: 617 IYGTFYVLTLRLVTSIIK 634
```

A related GBS gene <SEQ ID 8639> and protein <SEQ ID 8640> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:     -11.64
GvH: Signal Score (-7.5): -3.52
     Possible site: 37
>>> Seems to have no N-terminal signal sequence
ALOM program    count: 11 value: -11.62 threshold:  0.0
   INTEGRAL     Likelihood =-11.62   Transmembrane   55 -  71 (  49 -  75)
   INTEGRAL     Likelihood =-10.30   Transmembrane  197 - 213 ( 192 - 218)
   INTEGRAL     Likelihood = -9.13   Transmembrane  152 - 168 ( 141 - 172)
   INTEGRAL     Likelihood = -8.70   Transmembrane  624 - 640 ( 619 - 645)
   INTEGRAL     Likelihood = -8.44   Transmembrane  222 - 238 ( 219 - 250)
   INTEGRAL     Likelihood = -7.75   Transmembrane  283 - 299 ( 280 - 307)
```

```
   INTEGRAL    Likelihood = -7.70   Transmembrane  533 - 549 ( 526 - 552)
   INTEGRAL    Likelihood = -6.95   Transmembrane  108 - 124 (  99 - 140)
   INTEGRAL    Likelihood = -4.88   Transmembrane  585 - 601 ( 581 - 610)
   INTEGRAL    Likelihood = -3.82   Transmembrane   25 -  41 (  21 -  47)
   INTEGRAL    Likelihood = -0.48   Transmembrane  602 - 618 ( 602 - 618)
   PERIPHERAL  Likelihood =  1.16     129
  modified ALOM score:   2.82

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF02245(310 - 2262 of 2562)
GP|9802356|gb|AAF99695.1|AF267498_5|AF267498(1 - 639 of 640) permease OrfY {Streptococcus
mutans}
%Match = 10.2
%Identity = 24.0  %Similarity = 49.8
Matches = 147  Mismatches = 297  Conservative Sub.s = 158

123       153       183       213       243       273       303       333
QKTC*IYLKLLTWMDKLF*W*PIQQMLLVMPNAFYLSKMDVFFTNFIVVIRIIANSIKIFL*QCLPY*GVNNMFYLKIAW
                                                                        ||  ||::
                                                                        MFLPKISF

363       393       423       453       474       504       534       564
HNLKHSIDQYIPFLLASLLLYSLTCSTLLILMSAVGRDMGTAAT---VLFLGVIVLSIFAVVMEHYSYNILMKQRSSEFG
|||  :    :|::    ::  :    :  | :    :: ||     :| :| |::|:: ::   |:   :  :|:|  |
HNLIVNKSLTLPYFAIMTIFSGFNYVLINFLTNPSFYNIPTARILIDILIFGFILISLLMLLYGRYANRFISDERNSNMG
          20        30        40        50        60        70        80

594       624       654       684       714       744       774       804
LYNILGMNKRQVARVASLELFIIYIFLISIGSLFSAFFAKFIYLIFVNIINYHALNLSLSLWPFIICIVIFTGIFLTLEV
:: :||| |:|: ::  || : ::      | :|   ::|  :|   |:|    :    ||      :::   |:: : :
IFLMLGMGKKQLLKIIYLEKLYLFTGTFFGGLIFGFVYSKIFFLFIRNLIVIGDVREQYSLTAISWLLILTFFIYFIIYL
         100       110       120       130       140       150       160

834       864       894       924       954                1011      1041
PVIRHVHLSSPLSLFRKKQQGEKEPKGNLILAILALVAIAIAYTMALTSGKAPALAVIY-RFFFAVLLVIAGTYLFYISF
   | :   |   :|  | : :   | :: : :: | |: : |  ||||   | :   : || :|  ||  | :  : |
SEYRLLKRQSITVIFNSKAKRDNPRKTSVFVGLFGLFALLMGYHFALTS---PNVTTSFSRFIYAACLVTLGIFCTFSSG
         180       190       200       210          220       230       240

1071      1101      1131      1161      1191      1221      1251      1281
MTWYLKRLRQNKHYYYKSEHFVSTSQMIFRMKQNAVGLASITLLAVMALVTIATTVSLYSNTQNVVTGLFPKSVSLSIDN
:   |  ::: :  ||    ||  : :  |:: ||: ||:| :::   ||:::   |||    |:|    |: |
VIMLLTVIKKRRAIYYNQRRFVVIASLFHRIRSNALSLATICIFSTATLVSLSVLASLYLAKDNMVRLSSPRDV------
            260       270       280       290       300       310

1311      1341      1371      1401      1431      1461
SKGDAKNIFEEKILKKLGKSSKEAITYNQTMISMPVSQSSELNITSKNVKHVDITKTGFM--------------------
                             |::|    : : ::| :||  || :|    :
-----------------------------TVLSTTDIEPNLMDIATKN--HVTLTNRQNLKVSQSVYGNIKGSHLSVDPN
                           320       330         340       350       360

1464      1494      1524      1554      1584      1614      1641      1671
---------YLITQNDFRRLGHQLPKLKDNQVAYFVQKGDSRLKKINLLGNKFDVVKNLKEA-YVPETTNTYNPGLIIFA
          :|: : |    :   :||::::  :|  | :        | |:  || :|   ::     :  | : |
GGMANDYQITVISLDSFNASNNTHYRLKNHEILTYVSNGAAAPSSYTTNGVKLTNVKQIKRINFIFSPLRSMQPNFFIIT
         380       390       400       410       420       430       440

1698      1728      1758      1788      1818      1842      1872      1902
NNKQI-DNIRKAYLPYTKNINTFPKTFKAYLDLNSQEINSISKNDIIEVDG--KYVGNISTKQSFLKEGYQMFGGLLFTG
:|::|   | |  | :         |   |  :  :::|     | :|     ::  |: : :         :|||||| |
```

```
DNREIIQTILKEELTWG--------TMAGY-HVKGKKMNQKDFYDELETTNFRQFSANVVSIRQVKSMFNALFGGLLFVG
         460               470        480       490       500       510

1932      1962      1992      2022      2052      2082      2112      2142
FLLGISFLLGIALIVYYKQYSEGHEDKRSYRILQEVGMSKKLVKRTINSQIMIFFFQPLVVAVIHFGVAIPMLKQMLLVF
 ::|  | :  |: :||:| |||  |:  |: : ::||: | :: :|  ||   |  |:   |::::  |:|:| :::  |
IIFGTIFAILTAITIYYQQLSEGIRDRDDYKAMIKLGMTNKTIQDSIKVQINFVFILPIAFALLNLIFALPILYKIMTTF
        530       540       550       560       570       580       590

2172      2202      2232      2262      2292      2322      2352      2382
GVLNSTIVYVVSGLTVLAISIIYFIIYRITSRTYYHIIER*KGLVILPILLH**KPID*KICYTK*KKEISYYFRRGYVT
|  :: :     |  ::     |: |   ||: || :| :
GFNDAGLFLRAVGTCLIVYLFFYWFICHCTSKLYYRLISKK
        610       620       630       640
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 687

A DNA sequence (GBSx0729) was identified in *S.agalactiae* <SEQ ID 2117> which encodes the amino acid sequence <SEQ ID 2118>. This protein is predicted to be ABC transporter OrfX. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5121(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF99694 GB:AF267498 ABC transporter OrfX [Streptococcus mutans]
 Identities = 118/242 (48%), Positives = 175/242 (71%), Gaps = 1/242 (0%)

Query: 5   INHLEKVFRTRFSKEETRALQDVDFKVEQGEFIAIMGESGSGKTTLLNILATLEKPTNGQ 64
            ++HL+KV++T+       AL+D+ F V++GEFIAIMGESGSGK+TLLNILA ++ P++G
Sbjct: 6   VSHLKKVYKTQEGLTN-EALKDITFSVQEGEFIAIMGESGSGKSTLLNILACMDYPSSGH 64

Query: 65  VILNGEDITKIKEAKLASFRLKNLGFVFQDFNLLDTLSVRDNIYLPLVLDRKRYKEMDHR 124
            +I N   + K+K+ + A FR +++GF+FQ+FNLL+  + +DN+ +P+++   +    + R
Sbjct: 65  IIFNNYQLEKVKDEEAAVFRSRHIGFIFQNFNLLNIFNNKDNLLIPVIISGSKVNSYEKR 124

Query: 125 LSELSSHLRIDDLLDKRPFELSGGQKQRVAIARSLITNPQILLADEPTAALDYRNSEDLL 184
            L +L++ + I+ LL K P+ELSGGQ+QR+AIAR+LI NP ++LADEPT  LD + S+ +L
Sbjct: 125 LRDLAAVVGIESLLSKYPYELSGGQQQRLAIARALIMNPDLILADEPTGQLDSKTSQRIL 184

Query: 185 NLFETINLDGQTILMVTHSANAASHAKRVLFIKDGRIFHQLYRGNKNNSEFNKDISLTMS 244
            NL   IN   +TILMVTHS  AAS+A RVLFIKDG IF+QL RG K+   F   I +  +
Sbjct: 185 NLLSNINAKRKTILMVTHSPKAASYANRVLFIKDGVIFNQLVRGCKSREGFLDQIIMAQA 244

Query: 245 AI 246
            ++
Sbjct: 245 SL 246
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2119> which encodes the amino acid sequence <SEQ ID 2120>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
             bacterial cytoplasm --- Certainty=0.2131(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 91/222 (40%), Positives = 142/222 (62%), Gaps = 2/222 (0%)

Query: 2   LLEINHLEKVFRTRFSKEETRALQDVDFKVEQGEFIAIMGESGSGKTTLLNILATLEKPT 61
           LL +  + K +      EE   L+ +D +V +G+F+AIMG SGSGK+TL+NI+  L+KP
Sbjct: 1   LLNLKDIRKSYH--LGTEEFAILKGIDLEVNEGDFLAIMGPSGSGKSTLMNIIGCLDKPG 58

Query: 62  NGQVILNGEDITKIKEAKLASFRLKNLGFVFQDFNLLDTLSVRDNIYLPLVLDRKRYKEM 121
           +G   + G D++ + + +LA  R + +GFVFQ+FNL+  L+   N+ LPL       KE
Sbjct: 59  SGSYAIEGRDVSSLSDNELADLRNQKIGFVFQNFNLMPKLTACQNVELPLTYMNVPKKER 118

Query: 122 DHRLSELSSHLRIDDLLDKRPFELSGGQKQRVAIARSLITNPQILLADEPTAALDYRNSE 181
             R  E+   + +++  + +P ELSGGQKQRVAIAR+L+TNP  +L DEPT ALD + S
Sbjct: 119 RKRALEMLKLVGLEERSEFKPMELSGGQKQRVAIARALVTNPSFILGDEPTGALDTKTSV 178

Query: 182 DLLNLFETINLDGQTILMVTHSANAASHAKRVLFIKDGRIFH 223
            +++LF+  N +G+TI+++TH    A+  K+ + ++DG I H
Sbjct: 179 QIMDLFKQFNDNGKTIIIITHEPEVAALCKKTVILRDGNIEH 220
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 688

A DNA sequence (GBSx0730) was identified in *S.agalactiae* <SEQ ID 2121> which encodes the amino acid sequence <SEQ ID 2122>. This protein is predicted to be nisin-resistance protein. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.16    Transmembrane    8 -  24 (   1 -  31)

----- Final Results -----
              bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB08491 GB:U25181 nisin-resistance protein [Lactococcus lactis]
 Identities = 108/318 (33%), Positives = 190/318 (58%), Gaps = 8/318 (2%)

Query: 3   RKIVLLFVVPMLIVLGILGVVVHYYGSALNIYLLPPSSERYGRVILDRVEQRGLYSQGRQ 62
           ++I+L  V    + LGI    ++++G   NIYL+PPS ++Y RV L  +++ GL++  ++
Sbjct: 5   KRILLGLVAVCALFLGI----IYFWGYKFNIYLVPPSPQKYVRVALKNMDELGLFTDSKE 60

Query: 63  WQIIRQRSEKKLKTSKSYQESRNIVQEAVRYGGGKHSQILSKETVRRDTLDSRYPEYRRL 122
           W   ++++ ++   +K+Y E+   +Q+A++  GGKHS I  +E + + ++          +
Sbjct: 61  WVETKKKTIEETSNAKNYAETIPFLQKAIKVAGGKHSFIEHEEDISKRSITKYIKPKAEI 120

Query: 123 NEDILLITIPSISKLDKRSISHYSGKLQNILMEKSYKGLILDLSNNTGGNMIPMIGGVAS 182
             + L++TIP  +  D ++ S Y+  L++   + +Y G+I+DL  N GG++ PM+ G++
Sbjct: 121 EGNTLILTIPEFTGNDSQA-SDYANFLESSFHKNNYNGVIVDLRGNRGGDLSPMVLGLSP 179

Query: 183 ILPNDTLFHYTDKYGNKKTITMKNIPLEALKISRKTINTKHV---PIAIITNHKTASSAE 239
           +LP+ TLF Y DK  + K + ++N  + +   S K  + K +   PIA++ ++ T SS E
Sbjct: 180 LLPDGTLFTYVDKSSHSKPVELQNGEINSGGSSTKVSDNKKIKKAPIAVLIDNNTGSSGE 239

Query: 240 MTFLSFKGLPNVKSFGQATAGYTTVNETFMLYDGARLALTTGIVSDRQGYKYENTPILPD 299
           +T L FKG+PNVK  G  +AGYT+ N+T  LYDG+ L +T+  V DR    Y+N PI PD
```

```
Sbjct: 240 LTALCFKGIPNVKFLGSDSAGYTSANQTVYLYDGSTLQITSAFVKDRTNNIYKNFPISPD 299

Query: 300 QVTSLPLQESQSWLKSRI 317
             T+     +  W+KS+I
Sbjct: 300 IQTNNAKSSAIEWIKSQI 317
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8641> and protein <SEQ ID 8642> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:     12.71
GvH: Signal Score (-7.5): -5.64
     Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value: -13.16 threshold:  0.0
   INTEGRAL    Likelihood =-13.16   Transmembrane    8 -  24 (   1 -  31)
   PERIPHERAL  Likelihood =  4.03     174
 modified ALOM score:   3.13

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
34.7/62.5% over 311aa
                                       Lactococcus lactis
  GP|805128| nisin-resistance protein Insert characterized

ORF01108(343 - 1254 of 1560)
GP|805128|gb|AAB08491.1||U25181(7 - 318 of 318) nisin-resistance protein {Lactococcus
lactis}
%Match = 19.4
%Identity = 34.6  %Similarity = 62.4
Matches = 106  Mismatches = 112  Conservative Sub.s = 85

231       261       291       321       351                  393       423
LKLSNL*EIGLKM*GYSKPFCHIIDLKRKGEQEMRRKIVLLFVVPMLIVLGILGV------VVHYYGSALNIYLLPPSSE
                                           :   |:||:: |      :::::|  :||||:||| :
                                         MKIGKRILLGLVAVCALFLGIIYFWGYKFNIYLVPPSPQ
                                                  10        20        30

453       483       513       543       573       603       633       663
RYGRVILDRVEQRGLYSQGRQWQIIRQRSEKKLKTSKSYQESRNIVQEAVRYGGGKHSQILSKETVRRDTLDSRYPEYRR
:| || |  ::: ||::  ::|   :::: ::   :|:| |:   :|:|::  ||||| |  :| : : ::
KYVRVALKNMDELGLFTDSKEWVETKKKTIEETSNAKNYAETIPFLQKAIKVAGGKHSFIEHEEDISKRSITKYIKPKAE
         50        60        70        80        90       100       110

693       723       753       783       813       843       873       903
LNEDILLITIPSISKLDKRSISHYSGKLQNILMEKSYKGLILDLSNNTGGNMIPMIGGVASILPNDTLFHYTDKYGNKKT
:  : |::|||  :  | :: | |:  |:: : : :| |:|:||  | ||:: ||: |:: :||: ||| | ||  : |
IEGNTLILTIPEFTGNDSQA-SDYANFLESSFHKNNYNGVIVDLRGNRGGDLSPMVLGLSPLLPDGTLFTYVDKSSHSKP
        130       140        150       160       170       180       190

933       963       984       1014      1044      1074      1104      1134
ITMKNIPLEALKISRKTINTKHV---PIAIITNHKTASSAEMTFLSFKGLPNVKSFGQATAGYTTVNETFMLYDGARLAL
: ::|  : :   | |  : | :   |||:: :: | || |:| | |||:|||| :|  :||||: |:|  ||||: | :
VELQNGEINSGGSSTKVSDNKKIKKAPIAVLIDNNTGSSGELTALCFKGIPNVKFLGSDSAGYTSANQTVYLYDGSTLQI
         210       220       230       240       250       260       270

1164      1194      1224      1254      1284      1314      1344      1374
TTGIVSDRQGYKYENTPILPDQVTSLPLQESQSWLKSRINQN*GIINKGELYVIRNQSLRKSFSYTFFKRRDKGSTRRRF
```

```
|:  | ||    |:| || ||  |:     :  |:||:|
TSAFVKDRTNNIYKNFPISPDIQTNNAKSSAIEWIKSQIK
```

SEQ ID 2122 (GBS38) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 7; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 12; MW 62kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 689

A DNA sequence (GBSx0731) was identified in *S.agalactiae* <SEQ ID 2123> which encodes the amino acid sequence <SEQ ID 2124>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2125> which encodes the amino acid sequence <SEQ ID 2126>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1369(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 31/49 (63%), Positives = 43/49 (87%)

Query: 6  KKLTKSLGPIGKLISIIPDTTELIGKAIDNSRPIIEKELDRRHEKKTDL 54
           K++ K+LG +GKL+SI+PDTTE+IGK IDNSRPIIEK ++++HEK+  L
Sbjct: 3  KRIRKALGVVGKLMSIVPDTTEIIGKTIDNSRPIIEKRMEQKHEKEMQL 51
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 690

A DNA sequence (GBSx0732) was identified in *S.agalactiae* <SEQ ID 2127> which encodes the amino acid sequence <SEQ ID 2128>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3644(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 2126.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 691

A DNA sequence (GBSx0733) was identified in *S.agalactiae* <SEQ ID 2129> which encodes the amino acid sequence <SEQ ID 2130>. This protein is predicted to be 28 kd outer membrane protein precursor (yaeC). Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> May be a lipoprotein

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB59827 GB:AJ012388 hypothetical protein [Lactococcus lactis]
 Identities = 123/290 (42%), Positives = 178/290 (60%), Gaps = 18/290 (6%)

Query: 1   MKIKKLLGLTTTVVISALILGAC------GQSKNEDAKVVRVGTMVKSKTEKARWDKIEE 54
           +K +++L +T  +++  +I+G        G      +K+V++G M   K E   W ++++
Sbjct: 3   VKNRRIL-ITIIILVFIIIVGGIFAFSHSGNKSKVSSKIVKIGLMPGGKQEDVIWKQVQK 61

Query: 55  LVKKK-GVKLKFTEFTDYTQPNKALESDEIDINAFQHYNYLNNWNKANKTNLVSVAETYF 113
             K + G+ LKF  FTD  +PNKAL + E+D+NAFQHY YL +WNKAN  N+VS+ +T
Sbjct: 62  NAKDQFGITLKFVNFTDGDEPNKALVNHEVDLNAFQHYAYLKSWNKANNGNIVSIGDTII 121

Query: 114 TSFRLYSGTKNGKGKYQTVSEIPNKATITIPNDAVNESRSLYLLQSAGLLKLKVSGDALA 173
           T   LYS       KY+ V EIP+K+TI IPND  NESR+LY+L++AGL+KL  S   LA
Sbjct: 122 TPIHLYST------KYKKVDEIPDKSTIAIPNDITNESRALYVLKNAGLIKLDTSRGVLA 175

Query: 174 TMSDVVSNPKSLDLKEVDAAQTARSLDSTDAAVINNDFVTEAGINPKSAIFIEPKSKNAK 233
           T+ D+  NPKSL +KE+DA+QT R+LDS  AAVIN +F   A  + K +I+ EP ++++
Sbjct: 176 TVKDIRENPKSLIIKEIDASQTPRALDSVAAAVINYNFAISAKNSDKESIYQEPLNEDSA 235

Query: 234 QWYNLLVAQKGWQDKSKAKAIKEVVKAYHTDAVKKVIEKT-SQGLDQPVW 282
           QW N + A    Q     K  KEVVKAY    +  +I+K    G + P W
Sbjct: 236 QWINFIAAN---QSDKNNKVYKEVVKAYEQKNIADIIKKEYPDGGELPAW 282
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2131> which encodes the amino acid sequence <SEQ ID 2132>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1766(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 145/264 (54%), Positives = 203/264 (75%), Gaps = 2/264 (0%)

Query: 20  LGACGQSKNEDAKVVRVGTMVKSKTEKARWDKIEELVKKKGVKLKFTEFTDYTQPNKALE 79
           L AC + K +D   + +G M K+++++ARWDK+EEL+KK  + LK+ EFTDY+QPNKA+
Sbjct: 1   LVACSE-KQDDKNTLTIGVMTKTESDQARWDKVEELLKKDNITLKYKEFTDYSQPNKAVA 59
```

```
Query: 80  SDEIDINAFQHYNYLNNWNKANKTNLVSVAETYFTSFRLYSGT-KNGKGKYQTVSEIPNK 138
           + E+DINAFQHYN+LNNWNK NK +LV++A+TY +   L+SGT ++GK KY++V+++PN
Sbjct: 60  NGEVDINAFQHYNFLNNWNKENKEHLVAIADTYISPINLFSGTSQDGKAKYKSVADLPNG 119

Query: 139 ATITIPNDAVNESRSLYLLQSAGLLKLKVSGDALATMSDVVSNPKSLDLKEVDAAQTARS 198
             I +PNDA NESR+LY+LQSAGL+KL VSGD LAT++++  N K LD+KE+DA+QTAR+
Sbjct: 120 TQIAVPNDATNESRALYVLQSAGLIKLNVSGDQLATIANISENKKKLDIKELDASQTARA 179

Query: 199 LDSTDAAVINNDFVTEAGINPKSAIFIEPKSKNAKQWYNLLVAQKGWQDKSKAKAIKEVV 258
           L S DAAV+NN +   A I+ K+++F E    N+KQW N++  QK W+   KA AIK+++
Sbjct: 180 LVSADAAVVNNSYAVPAKIDYKTSLFKEKADDNSKQWINIIAGQKDWEKSEKADAIKKLI 239

Query: 259 KAYHTDAVKKVIEKTSQGLDQPVW 282
           KAY TD VKKV+EKTS G+D  VW
Sbjct: 240 KAYQTDEVKKVVEKTSNGIDVSVW 263
```

SEQ ID 2130 (GBS96) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 7; MW 32kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 22 (lane 3; MW 57.2kDa).

The GBS96-GST fusion product was purified (Figure 195, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 290), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 692

A DNA sequence (GBSx0734) was identified in *S.agalactiae* <SEQ ID 2133> which encodes the amino acid sequence <SEQ ID 2134>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5103(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9807> which encodes amino acid sequence <SEQ ID 9808> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes.*

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 693

A DNA sequence (GBSx0735) was identified in *S.agalactiae* <SEQ ID 2135> which encodes the amino acid sequence <SEQ ID 2136>. This protein is predicted to be glucose-inhibited division protein (gid). Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0656(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13486 GB:Z99112 glucose-inhibited division protein [Bacillus subtilis]
 Identities = 289/439 (65%), Positives = 352/439 (79%), Gaps = 10/439 (2%)

Query: 1   MSQSYINVIGAGLAGSEAAYQIAKRGIPVKLYEMRGVKSTPQHKTDNFAELVCSNSFRGD 60
           M+Q  +NVIGAGLAGSEAA+Q+AKRGI VKLYEMR VK TP H TD FAELVCSNS R +
Sbjct: 1   MNQQTVNVIGAGLAGSEAAWQLAKRGIQVKLYEMRPVKQTPAHHTDKFAELVCSNSLRSN 60

Query: 61  SLTNAVGLLKEEMRRLDSIIMRNGEAHRVPAGGAMAVDREGYSEAVTEEIHKHPLIEVIR 120
           +L NAVG+LKEEMR LDS I+   +   VPAGGA+AVDR  ++ +VT  +  HP + VI
Sbjct: 61  TLANAVGVLKEEMRALDSAIIAAADECSVPAGGALAVDRHEFAASVTNRVKNHPNVTVIN 120

Query: 121 DEITDIPGDAITVIATGPLTSDSLAAKIHELNGGDGFYFYDAAAPIVDKNTIDINKVYLK 180
           +E+T+IP +  T+IATGPLTS+SL+A++ EL G D  YFYDAAAPIV+K+++D++KVYLK
Sbjct: 121 EEVTEIP-EGPTIIATGPLTSESLSAQLKELTGEDYLYFYDAAAPIVEKDSLDMDKVYLK 179

Query: 181 SRYDKGEAAYLNCPMTKEEFMAFHEALTTAEEAPLNSFEKEKYFEGCMPIEVMAKRGIKT 240
           SRYDKGEAAYLNCPMT+EEF  FHEALT+AE  PL  FEKE +FEGCMPIEVMAKRG KT
Sbjct: 180 SRYDKGEAAYLNCPMTEEEFDRFHEALTSAETVPLKEFEKEIFFEGCMPIEVMAKRGKKT 239

Query: 241 MLYGPMKPVGLEYPEDYKGPRDGEFKTPYAVVQLRQDNAAGSLYNIVGFQTHLKWGEQKR 300
           ML+GPMKPVGLE+P   K         PYAVVQLRQD+AAG+LYNIVGFQTHLKWG+QK
Sbjct: 240 MLFGPMKPVGLEHPVTGK--------RPYAVVQLRQDDAAGTLYNIVGFQTHLKWGDQKE 291

Query: 301 VFQMIPGLENAEFVRYGVMHRNSYMDSPNLLNQTFATRKNPNLFFAGQMTGVEGYVESAA 360
           V ++IPGLEN E VRYGVMHRN++++SP+LL  T+  +   +LFFAGQMTGVEGYVESAA
Sbjct: 292 VLKLIPGLENVEIVRYGVMHRNTFINSPSLLKPTYQFKNRSDLFFAGQMTGVEGYVESAA 351

Query: 361 SGLVAGINAVRRFNGESEVVFPQTTAIGALPHYITHTDSKHFQPMNVNFGIIKELEGPRI 420
           SGLVAGINA +   GE  V+FPQ TAIG++ HYIT T+ K+FQPMN NFG++KEL   +I
Sbjct: 352 SGLVAGINAAKLVLGEELVIFPQETAIGSMAHYITTTNQKNFQPMNANFGLLKELP-VKI 410

Query: 421 RDKKERYEAIATRALKDLE 439
           ++KKER E  A RA++ ++
Sbjct: 411 KNKKERNEQYANRAIETIQ 429
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2137> which encodes the amino acid sequence <SEQ ID 2138>. Analysis of this protein sequence reveals the following:

```
      Possible site: 30

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.44   Transmembrane   12 -  28 (   9 -  32)

----- Final Results -----
               bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 111-113
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB13486 GB:Z99112 glucose-inhibited division protein [Bacillus subtilis]
 Identities = 292/435 (67%), Positives = 350/435 (80%), Gaps = 10/435 (2%)

Query: 59  INVIGAGLAGSEAAYQIAKRGIPVKLYEMRGVKATPQHKTTNFAELVCSNSFRGDSLTNA 118
           +NVIGAGLAGSEAA+Q+AKRGI VKLYEMR VK TP H T  FAELVCSNS R ++L NA
Sbjct: 6   VNVIGAGLAGSEAAWQLAKRGIQVKLYEMRPVKQTPAHHTDKFAELVCSNSLRSNTLANA 65

Query: 119 VGLLKEEMRRLDSIIMRNGEANRVPAGGAMAVDREGYAESVTAELENHPLIEVIRGEITE 178
```

```
           VG+LKEEMR LDS I+    +  VPAGGA+AVDR  +A SVT  ++NHP + VI  E+TE
Sbjct: 66  VGVLKEEMRALDSAIIAAADECSVPAGGALAVDRHEFAASVTNRVKNHPNVTVINEEVTE 125

Query: 179 IPDDAITVIATGPLTSDALAEKIHALNGGDGFYFYDAAAPIIDKSTIDMSKVYLKSRYDK 238
           IP+   T+IATGPLTS++L+ ++  L G D  YFYDAAAPI++K ++DM KVYLKSRYDK
Sbjct: 126 IPEGP-TIIATGPLTSESLSAQLKELTGEDYLYFYDAAAPIVEKDSLDMDKVYLKSRYDK 184

Query: 239 GEAAYLNCPMTKEEFMAFHEALTTAEEAPLNAFEKEKYFEGCMPIEVMAKRGIKTMLYGP 298
           GEAAYLNCPMT+EEF  FHEALT+AE  PL  FEKE +FEGCMPIEVMAKRG KTML+GP
Sbjct: 185 GEAAYLNCPMTEEEFDRFHEALTSAETVPLKEFEKEIFFEGCMPIEVMAKRGKKTMLFGP 244

Query: 299 MKPVGLEYPDDYTGPRDGEFKTPYAVVQLRQDNAAGSLYNIVGFQTHLKWGEQKRVFQMI 358
           MKPVGLE+P   TG R      PYAVVQLRQD+AAG+LYNIVGFQTHLKWG+QK V ++I
Sbjct: 245 MKPVGLEHP--VTGKR------PYAVVQLRQDDAAGTLYNIVGFQTHLKWGDQKEVLKLI 296

Query: 359 PGLENAEFVRYGVMHRNSYMDSPNLLTETFQSRSNPNLFFAGQMTGVEGYVESAASGLVA 418
           PGLEN E VRYGVMHRN++++SP+LL  T+Q ++  +LFFAGQMTGVEGYVESAASGLVA
Sbjct: 297 PGLENVEIVRYGVMHRNTFINSPSLLKPTYQFKNRSDLFFAGQMTGVEGYVESAASGLVA 356

Query: 419 GINAARLFKREEALIFPQTTAIGSLPHYVTHADSKHFQPMNVNFGIIKELEGPRIRDKKE 478
           GINAA+L   EE +IFPQ TAIGS+ HY+T  + K+FQPMN NFG++KEL   +I++KKE
Sbjct: 357 GINAAKLVLGEELVIFPQETAIGSMAHYITTTNQKNFQPMNANFGLLKELP-VKIKNKKE 415

Query: 479 RYEAIASRALADLDT 493
           R E  A+RA+  + T
Sbjct: 416 RNEQYANRAIETIQT 430
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 395/439 (89%), Positives = 417/439 (94%)

Query: 4   SYINVIGAGLAGSEAAYQIAKRGIPVKLYEMRGVKSTPQHKTDNFAELVCSNSFRGDSLT 63
           +YINVIGAGLAGSEAAYQIAKRGIPVKLYEMRGVK+TPQHKT NFAELVCSNSFRGDSLT
Sbjct: 57  TYINVIGAGLAGSEAAYQIAKRGIPVKLYEMRGVKATPQHKTTNFAELVCSNSFRGDSLT 116

Query: 64  NAVGLLKEEMRRLDSIIMRNGEAHRVPAGGAMAVDREGYSEAVTEEIHKHPLIEVIRDEI 123
           NAVGLLKEEMRRLDSIIMRNGEA+RVPAGGAMAVDREGY+E+VT E+  HPLIEVIR EI
Sbjct: 117 NAVGLLKEEMRRLDSIIMRNGEANRVPAGGAMAVDREGYAESVTAELENHPLIEVIRGEI 176

Query: 124 TDIPGDAITVIATGPLTSDSLAAKIHELNGGDGFYFYDAAAPIVDKNTIDINKVYLKSRY 183
           T+IP DAITVIATGPLTSD+LA KIH LNGGDGFYFYDAAAPI+DK+TID++KVYLKSRY
Sbjct: 177 TEIPDDAITVIATGPLTSDALAEKIHALNGGDGFYFYDAAAPIIDKSTIDMSKVYLKSRY 236

Query: 184 DKGEAAYLNCPMTKEEFMAFHEALTTAEEAPLNSFEKEKYFEGCMPIEVMAKRGIKTMLY 243
           DKGEAAYLNCPMTKEEFMAFHEALTTAEEAPLN+FEKEKYFEGCMPIEVMAKRGIKTMLY
Sbjct: 237 DKGEAAYLNCPMTKEEFMAFHEALTTAEEAPLNAFEKEKYFEGCMPIEVMAKRGIKTMLY 296

Query: 244 GPMKPVGLEYPEDYKGPRDGEFKTPYAVVQLRQDNAAGSLYNIVGFQTHLKWGEQKRVFQ 303
           GPMKPVGLEYP+DY GPRDGEFKTPYAVVQLRQDNAAGSLYNIVGFQTHLKWGEQKRVFQ
Sbjct: 297 GPMKPVGLEYPDDYTGPRDGEFKTPYAVVQLRQDNAAGSLYNIVGFQTHLKWGEQKRVFQ 356

Query: 304 MIPGLENAEFVRYGVMHRNSYMDSPNLLNQTFATRKNPNLFFAGQMTGVEGYVESAASGL 363
           MIPGLENAEFVRYGVMHRNSYMDSPNLL +TF +R NPNLFFAGQMTGVEGYVESAASGL
Sbjct: 357 MIPGLENAEFVRYGVMHRNSYMDSPNLLTETFQSRSNPNLFFAGQMTGVEGYVESAASGL 416

Query: 364 VAGINAVRRFNGESEVVFPQTTAIGALPHYITHTDSKHFQPMNVNFGIIKELEGPRIRDK 423
           VAGINA R F  E  ++FPQTTAIG+LPHY+TH DSKHFQPMNVNFGIIKELEGPRIRDK
Sbjct: 417 VAGINAARLFKREEALIFPQTTAIGSLPHYVTHADSKHFQPMNVNFGIIKELEGPRIRDK 476

Query: 424 KERYEAIATRALKDLEKFL 442
           KERYEAIA+RAL DL+  L
Sbjct: 477 KERYEAIASRALADLDTCL 495
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 694

A DNA sequence (GBSx0736) was identified in *S.agalactiae* <SEQ ID 2139> which encodes the amino acid sequence <SEQ ID 2140>. This protein is predicted to be transcriptional regulator (GntRfamily). Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5103(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04138 GB:AP001508 transcriptional regulator (GntR family)
           [Bacillus halodurans]
 Identities = 83/229 (36%), Positives = 133/229 (57%), Gaps = 1/229 (0%)

Query: 2   LPAYIKIHDAIKKEIDKGTWKIGQRLPSERDLADDYSVSRMTLRQSITLLVEEGILERRV 61
            LP Y +I + IK++I+ G  K G  L SER+ A+ Y VSRMT+RQ+I  LV +G + ++
Sbjct: 8   LPIYYQIEEQIKQQIESGVLKPGDMLKSEREYAEYYDVSRMTVRQAINNLVNQGYIYKKK 67

Query: 62  GSGTYVASHRVQEKMRGTTSFTEIVNSQGRKPSSKLISFQRKLANETEIQKLNLSQSDYV 121
            GSGTYV   ++++ + G TSFTE +  +G +PSS+L+ F+   A     ++LNL ++  V
Sbjct: 68  GSGTYVQEKKIEQALNGLTSFTEDMRKRGMEPSSRLLKFELIPATAKIAKELNLKENTPV 127

Query: 122 VRMERVRYADKVPLVYEVASIPENLIKGFEQSEVTEHFFKTLTEN-GYEIGKSQQTIYAR 180
              ++R+RY D VP+  E   +P NL+KG  +  + +  ++ + E     I  + Q I A
Sbjct: 128 TEIKRIRYGDGVPIAIERNLLPANLVKGLNEEIINQSLYQYIEEELNLRIADALQVIEAS 187

Query: 181 NASERVASHLEVNAGHAILALTQVSYFTDGKPFEYVHGQYVGDRFEFYL 229
             AS+  A  LE+  G  IL + + ++  DG   E V   Y  DR++F +
Sbjct: 188 TASKTEADLLEIQKGSPILLIERKTFLADGTVLELVKSAYRADRYKFMI 236
```

There is also homology to SEQ ID 1256.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 695

A DNA sequence (GBSx0737) was identified in *S.agalactiae* <SEQ ID 2141> which encodes the amino acid sequence <SEQ ID 2142>. This protein is predicted to be GMP synthase (guaA). Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.96   Transmembrane  228 - 244 ( 228 - 245)

----- Final Results -----
               bacterial membrane --- Certainty=0.1383(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD15805 GB:AF058326 GMP synthase [Lactococcus lactis]
 Identities = 416/511 (81%), Positives = 467/511 (90%), Gaps = 3/511 (0%)

Query: 10  IQKIIVLDYGSQYNQLIARRIREFGVFSELKSHKITADEIRDINPIGIVLSGGPNSVYAD 69
            ++KIIVLDYGSQYNQLIARRIRE GVFSEL SHK+TA EIR+INPIGI+LSGGPNSVY +
Sbjct: 6   LEKIIVLDYGSQYNQLIARRIREIGVFSELMSHKVTAKEIREINPIGIILSGGPNSVYDE 65
```

```
Query: 70  GAFGIDEEIFELGIPILGICYGMQLITHKLGGKVLPAGEAGHREYGQSALRLRSESALFA 129
           G+F ID EIFELG+P+LGICYGMQL+++KLGG V  AGE   REYG + L+L  +SALFA
Sbjct: 66  GSFDIDPEIFELGLPVLGICYGMQLMSYKLGGMVEAAGE---REYGVAPLQLTEKSALFA 122

Query: 130 GTPQEQLVLMSHGDAVTEIPEGFHLVGDSVDCPFAAMENTEKQFYGIQFHPEVRHSVYGN 189
           GTP+ Q VLMSHGD VT IPEGFH+VG S + PFAA+ENTE+  YGIQFHPEVRHSV+G
Sbjct: 123 GTPEVQDVLMSHGDRVTAIPEGFHVVGTSPNSPFAAVENTERNLYGIQFHPEVRHSVHGT 182

Query: 190 DILKNFAVNICGARGDWSMDNFIDMEIAKIRETVGDRKVLLGLSGGVDSSVVGVLLQRAI 249
           ++L+NFA+NICGA+G+WSM+NFIDM+I  IRE VGD+KVLLGLSGGVDSSVVGVLLQRAI
Sbjct: 183 EMLRNFALNICGAKGNWSMENFIDMQIKDIREKVGDKKVLLGLSGGVDSSVVGVLLQRAI 242

Query: 250 GDQLTCIFVDHGLLRKNEGDQVMDMLGGKFGLNIIRVDASKRFLDLLSGVEDPERKRKII 309
           GDQLT IFVDHG LRK E DQVM+ LGGKFGLNII+VDA KRF+D L G+ DPE +RKII
Sbjct: 243 GDQLTSIFVDHGFLRKGEADQVMETLGGKFGLNIIKVDAQKRFMDKLVGLSDPETQRKII 302

Query: 310 GNEFVYVFDDEASKLKGVDFLAQGTLYTDIIESGTETAQTIKSHHNVGGLPEDMQFELIE 369
           GNEFVYVFDDEA+KL+GVDFLAQGTLYTD+IESGT+TAQTIKSHHNVGGLPEDMQF+LIE
Sbjct: 303 GNEFVYVFDDEANKLEGVDFLAQGTLYTDVIESGTDTAQTIKSHHNVGGLPEDMQFQLIE 362

Query: 370 PLNTLFKDEVRALGTALGMPDEVVWRQPFPGPGLAIRVMGEITEEKLETVRESDAILREE 429
           PLNTLFKDEVRALGT LGMPDE+VWRQPFPGPGLAIRV+G++TEEKLETVRESDAILREE
Sbjct: 363 PLNTLFKDEVRALGTQLGMPDEIVWRQPFPGPGLAIRVLGDLTEEKLETVRESDAILREE 422

Query: 430 IAKAGLDRDVWQYFTVNTGVRSVGVMGDGRTYDYTIAIRAITSIDGMTADFAQLPWDVLK 489
           IA +GL+RDVWQYFTVNT V+SVGVMGD RTYDYT+AIRAITSIDGMTADFAQLPWD+L+
Sbjct: 423 IAASGLERDVWQYFTVNTDVKSVGVMGDQRTYDYTLAIRAITSIDGMTADFAQLPWDLLQ 482

Query: 490 KISTRIVNEVDHVNRIVYDITSKPPATVEWE 520
           KIS RIVNEVDHVNRIVYDITSKPPATVEW+
Sbjct: 483 KISKRIVNEVDHVNRIVYDITSKPPATVEWQ 513
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2143> which encodes the amino acid sequence <SEQ ID 2144>. Analysis of this protein sequence reveals the following:

```
     Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.96   Transmembrane  228 - 244 ( 228 - 245)

----- Final Results -----
               bacterial membrane --- Certainty=0.1383(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 203-205
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD15805 GB:AF058326 GMP synthase [Lactococcus lactis]
 Identities = 411/511 (80%), Positives = 464/511 (90%), Gaps = 3/511 (0%)

Query: 10  VQKIIVLDYGSQYNQLIARRIREFGVFSELKSHKITAQELREINPIGIVLSGGPNSVYAD 69
           ++KIIVLDYGSQYNQLIARRIRE GVFSEL SHK+TA+E+REINPIGI+LSGGPNSVY +
Sbjct: 6   LEKIIVLDYGSQYNQLIARRIREIGVFSELMSHKVTAKEIREINPIGIILSGGPNSVYDE 65

Query: 70  NAFGIDPEIFELGIPILGICYGMQLITHKLGGKVVPAGQAGNREYGQSTLHLRETSKLFS 129
            +F IDPEIFELG+P+LGICYGMQL+++KLGG V  AG+   REYG + L L E S LF+
Sbjct: 66  GSFDIDPEIFELGLPVLGICYGMQLMSYKLGGMVEAAGE---REYGVAPLQLTEKSALFA 122

Query: 130 GTPQEQLVLMSHGDAVTEIPEGFHLVGDSNDCPYAAIENTEKNLYGIQFHPEVRHSVYGN 189
           GTP+ Q VLMSHGD VT IPEGFH+VG S + P+AA+ENTE+NLYGIQFHPEVRHSV+G
Sbjct: 123 GTPEVQDVLMSHGDRVTAIPEGFHVVGTSPNSPFAAVENTERNLYGIQFHPEVRHSVHGT 182

Query: 190 DILKNFAISICGARGDWSMDNFIDMEIAKIRETVGDRKVLLGLSGGVDSSVVGVLLQKAI 249
           ++L+NFA++ICGA+G+WSM+NFIDM+I  IRE VGD+KVLLGLSGGVDSSVVGVLLQ+AI
Sbjct: 183 EMLRNFALNICGAKGNWSMENFIDMQIKDIREKVGDKKVLLGLSGGVDSSVVGVLLQRAI 242
```

```
Query: 250 GDQLTCIFVDHGLLRKDEGDQVMGMLGGKFGLNIIRVDASKRFLDLLADVEDPEKKRKII 309
           GDQLT IFVDHG LRK E DQVM  LGGKFGLNII+VDA KRF+D L  + DPE +RKII
Sbjct: 243 GDQLTSIFVDHGFLRKGEADQVMETLGGKFGLNIIKVDAQKRFMDKLVGLSDPETQRKII 302

Query: 310 GNEFVYVFDDEASKLKGVDFLAQGTLYTDIIESGTETAQTIKSHHNVGGLPEDMQFELIE 369
           GNEFVYVFDDEA+KL+GVDFLAQGTLYTD+IESGT+TAQTIKSHHNVGGLPEDMQF+LIE
Sbjct: 303 GNEFVYVFDDEANKLEGVDFLAQGTLYTDVIESGTDTAQTIKSHHNVGGLPEDMQFQLIE 362

Query: 370 PLNTLFKDEVRALGIALGMPEEIVWRQPFPGPGLAIRVMGAITEEKLETVRESDAILREE 429
           PLNTLFKDEVRALG  LGMP+EIVWRQPFPGPGLAIRV+G +TEEKLETVRESDAILREE
Sbjct: 363 PLNTLFKDEVRALGTQLGMPDEIVWRQPFPGPGLAIRVLGDLTEEKLETVRESDAILREE 422

Query: 430 IAKAGLDRDVWQYFTVNTGVRSVGVMGDGRTYDYTIAIRAITSIDGMTADFAQLPWDVLK 489
           IA +GL+RDVWQYFTVNT V+SVGVMGD RTYDYT+AIRAITSIDGMTADFAQLPWD+L+
Sbjct: 423 IAASGLERDVWQYFTVNTDVKSVGVMGDQRTYDYTLAIRAITSIDGMTADFAQLPWDLLQ 482

Query: 490 KISTRIVNEVDHVNRIVYDITSKPPATVEWE 520
           KIS RIVNEVDHVNRIVYDITSKPPATVEW+
Sbjct: 483 KISKRIVNEVDHVNRIVYDITSKPPATVEWQ 513
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 487/520 (93%), Positives = 505/520 (96%)

Query: 1   MTDISILNDIQKIIVLDYGSQYNQLIARRIREFGVFSELKSHKITADEIRDINPIGIVLS 60
           MT+ISILND+QKIIVLDYGSQYNQLIARRIREFGVFSELKSHKITA E+R+INPIGIVLS
Sbjct: 1   MTEISILNDVQKIIVLDYGSQYNQLIARRIREFGVFSELKSHKITAQELREINPIGIVLS 60

Query: 61  GGPNSVYADGAFGIDEEIFELGIPILGICYGMQLITHKLGGKVLPAGEAGHREYGQSALR 120
           GGPNSVYAD AFGID EIFELGIPILGICYGMQLITHKLGGKV+PAG+AG+REYGQS L
Sbjct: 61  GGPNSVYADNAFGIDPEIFELGIPILGICYGMQLITHKLGGKVVPAGQAGNREYGQSTLH 120

Query: 121 LRSESALFAGTPQEQLVLMSHGDAVTEIPEGFHLVGDSVDCPFAAMENTEKQFYGIQFHP 180
           LR  S LF+GTPQEQLVLMSHGDAVTEIPEGFHLVGDS DCP+AA+ENTEK  YGIQFHP
Sbjct: 121 LRETSKLFSGTPQEQLVLMSHGDAVTEIPEGFHLVGDSNDCPYAAIENTEKNLYGIQFHP 180

Query: 181 EVRHSVYGNDILKNFAVNICGARGDWSMDNFIDMEIAKIRETVGDRKVLLGLSGGVDSSV 240
           EVRHSVYGNDILKNFA++ICGARGDWSMDNFIDMEIAKIRETVGDRKVLLGLSGGVDSSV
Sbjct: 181 EVRHSVYGNDILKNFAISICGARGDWSMDNFIDMEIAKIRETVGDRKVLLGLSGGVDSSV 240

Query: 241 VGVLLQRAIGDQLTCIFVDHGLLRKNEGDQVMDMLGGKFGLNIIRVDASKRFLDLLSGVE 300
           VGVLLQ+AIGDQLTCIFVDHGLLRK+EGDQVM MLGGKFGLNIIRVDASKRFLDLL+ VE
Sbjct: 241 VGVLLQKAIGDQLTCIFVDHGLLRKDEGDQVMGMLGGKFGLNIIRVDASKRFLDLLADVE 300

Query: 301 DPERKRKIIGNEFVYVFDDEASKLKGVDFLAQGTLYTDIIESGTETAQTIKSHHNVGGLP 360
           DPE+KRKIIGNEFVYVFDDEASKLKGVDFLAQGTLYTDIIESGTETAQTIKSHHNVGGLP
Sbjct: 301 DPEKKRKIIGNEFVYVFDDEASKLKGVDFLAQGTLYTDIIESGTETAQTIKSHHNVGGLP 360

Query: 361 EDMQFELIEPLNTLFKDEVRALGTALGMPDEVVWRQPFPGPGLAIRVMGEITEEKLETVR 420
           EDMQFELIEPLNTLFKDEVRALG ALGMP+E+VWRQPFPGPGLAIRVMG ITEEKLETVR
Sbjct: 361 EDMQFELIEPLNTLFKDEVRALGIALGMPEEIVWRQPFPGPGLAIRVMGAITEEKLETVR 420

Query: 421 ESDAILREEIAKAGLDRDVWQYFTVNTGVRSVGVMGDGRTYDYTIAIRAITSIDGMTADF 480
           ESDAILREEIAKAGLDRDVWQYFTVNTGVRSVGVMGDGRTYDYTIAIRAITSIDGMTADF
Sbjct: 421 ESDAILREEIAKAGLDRDVWQYFTVNTGVRSVGVMGDGRTYDYTIAIRAITSIDGMTADF 480

Query: 481 AQLPWDVLKKISTRIVNEVDHVNRIVYDITSKPPATVEWE 520
           AQLPWDVLKKISTRIVNEVDHVNRIVYDITSKPPATVEWE
Sbjct: 481 AQLPWDVLKKISTRIVNEVDHVNRIVYDITSKPPATVEWE 520
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 696

A DNA sequence (GBSx0740) was identified in *S.agalactiae* <SEQ ID 2145> which encodes the amino acid sequence <SEQ ID 2146>. This protein is predicted to be branched chain amino acid ABC transporter, periplasmic amino acid-bind. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0957(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9409> which encodes amino acid sequence <SEQ ID 9410> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD36211 GB:AE001771 branched chain amino acid ABC transporter,
           periplasmic amino acid-binding protein [Thermotoga maritima]
 Identities = 31/92 (33%), Positives = 51/92 (54%), Gaps = 4/92 (4%)

Query: 26  AKAFHDHYVKAYGEEPSMFSALSYDAVYMAAKSAKGAKTSID---IKKALAKLKDFKGVT 82
           AK F + Y + YG+EP+  +AL YDA YM    A     S D   I + + K ++F G +
Sbjct: 275 AKKFVEVYKEKYGKEPAALNALGYDA-YMVLLDAIERAGSFDREKIAEEIRKTRNFNGAS 333

Query: 83  GKMSIDKNHNVVKSAYVVKLEDGKTSSVNIIS 114
           G ++ID+N + +KS  V  +++G      +I+
Sbjct: 334 GIINIDENGDAIKSVVVNIVKNGSVDFEAVIN 365
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 9410 (GBS660) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 135 (lane 8 & 9; MW 71.5kDa) + lane 10; MW 27kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 141 (lane 2; MW 46.5kDa) and in Figure 181 (lane 3; MW 46kDa).

GBS660-His was purified as shown in Figure 233, lane 5-6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 697

A DNA sequence (GBSx0741) was identified in *S.agalactiae* <SEQ ID 2147> which encodes the amino acid sequence <SEQ ID 2148>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-10.61   Transmembrane  140 - 156 ( 129 - 158)
   INTEGRAL    Likelihood = -9.55   Transmembrane   60 -  76 (  53 -  80)
   INTEGRAL    Likelihood = -7.59   Transmembrane  264 - 280 ( 257 - 285)
   INTEGRAL    Likelihood = -5.79   Transmembrane  232 - 248 ( 219 - 251)
   INTEGRAL    Likelihood = -2.23   Transmembrane  190 - 206 ( 190 - 207)
   INTEGRAL    Likelihood = -1.75   Transmembrane   90 - 106 (  90 - 110)

----- Final Results -----
                bacterial membrane --- Certainty=0.5246(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10059> which encodes amino acid sequence <SEQ ID 10060> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD36212 GB:AE001771 branched chain amino acid ABC transporter,
           permease protein [Thermotoga maritima]
 Identities = 140/295 (47%), Positives = 200/295 (67%), Gaps = 7/295 (2%)

Query: 2   LQQLVNGLILGSIYALLALGYTMVYGIIKLINFAHGDIYMMGAFMGYYLINHLHLNFFLA 61
           LQ L NG++LG +YAL+A+GYTMVYGI++LINFAHGD+ MMG +  +Y    L LN   +
Sbjct: 5   LQNLFNGIMLGGLYALIAIGYTMVYGILRLINFAHGDVMMMGVYFAFYAATLLSLNPLFS 64

Query: 62  LLIAMLGSAFLGVVIEYLAYRPLRKSTRIAALITAIGVSFLLEYGMVYLVGADTRAFPQA 121
            ++A+LG+A LG +I+ +AY+PLR + RI+ALITAIGVSF LE   V + GA  ++F +
Sbjct: 65  AIVAILGAALLGFLIDRVAYKPLRNAPRISALITAIGVSFFLESLAVVVFGAIPKSFLKV 124

Query: 122 IHTVKYNLGPITITNVQL-----IILGIALLLMLTLQFIVQKTKMGKAMRALSVDSDAAQ 176
                      +T+   ++     +++ I  ++++ L FIV +TK+G AMRA+S+D
Sbjct: 125 FKDRTILNKVLTVAGARIPLLTFLVIFITAVILIVLFFIVYRTKIGMAMRAISMDIPTTA 184

Query: 177 LMGINVNRTISFTFALGSALAGAGGVLIGLYYNSVQPLMGVTPGLKAFVAAVLGGIGIIP 236
           LMG+NV+  I FTFALGSALA A G++  + + +V P MG  PGLKAF+AAV GGIG IP
Sbjct: 185 LMGVNVDAVIGFTFALGSALAAASGIMWAMRFPNVHPYMGFMPGLKAFIAAVFGGIGSIP 244

Query: 237 GAAIGGFVIGILETLATAL--GVSDFRDGIVYAILILIFLIRPAGILGKNIKEKV 289
           GA +GG ++G++E    A    V  +RD   + ILI+I L++P+G+LGK I EKV
Sbjct: 245 GAVLGGVLLGLIEIFLAAYFPAVMGYRDAFAFIILIIILLVKPSGLLGKKIVEKV 299
```

There is also homology to SEQ ID 2150. A related sequence was also identified in GAS <SEQ ID 9171> which encodes the amino acid sequence <SEQ ID 9172>. Analysis of this protein sequence reveals the following:

```
     Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL    Likelihood =-12.74   Transmembrane  196 - 212 ( 191 - 219)
     INTEGRAL    Likelihood =-12.42   Transmembrane   12 -  28 (   5 -  36)
     INTEGRAL    Likelihood = -7.22   Transmembrane  106 - 122 ( 102 - 126)
     INTEGRAL    Likelihood = -4.78   Transmembrane  242 - 258 ( 240 - 260)
     INTEGRAL    Likelihood = -2.50   Transmembrane   61 -  77 (  60 -  77)
     INTEGRAL    Likelihood = -2.34   Transmembrane  293 - 309 ( 291 - 309)
     INTEGRAL    Likelihood = -1.44   Transmembrane  139 - 155 ( 138 - 156)
     INTEGRAL    Likelihood = -1.33   Transmembrane  317 - 333 ( 317 - 333)

----- Final Results -----
               bacterial membrane --- Certainty=0.609(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 35/147 (23%), Positives = 71/147 (47%), Gaps = 6/147 (4%)

Query: 134 ITNVQLIILGI--ALLLMLTLQFIVQKTKMGKAMRALSVDSDAAQLMGINVNRTISFTFA 191
           +TN   I +GI  A++ +  + F++ KT +G  +R++ ++  A++  G++  RTI   +
Sbjct: 197 LTNNSRINIGIFFAIIAIALIWFLLNKTTLGFEIRSVGLNPHASEYAGMSSKRTIILSMI 256

Query: 192 LGSALAGAGGVL--IGLYYNSVQPLMGVTPGLKAFVAAVLGGIGIIPGAAIGGFVIGILE 249
           +  ALAG GGV+  +G + N       +  G      ++L     + G     F+ G+L
Sbjct: 257 ISGALAGLGGVVEGLGTFENVFVQGSSLAVGFDGMAVSLLAANSPL-GIFFSSFLFGVLN 315

Query: 250 TLATALGVSDFRDGIVYAILI-LIFLI 275
             A  + ++     +V  +   +IF +
Sbjct: 316 IGAPGMNIAGIPPELVKVVTASIIFFV 342
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 698

A DNA sequence (GBSx0742) was identified in *S.agalactiae* <SEQ ID 2151> which encodes the amino acid sequence <SEQ ID 2152>. This protein is predicted to be branched chain amino acid ABC transporter, permease protein (livM). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.76   Transmembrane   90 - 106 (  84 - 113)
   INTEGRAL    Likelihood = -8.23   Transmembrane   12 -  28 (   5 -  33)
   INTEGRAL    Likelihood = -8.17   Transmembrane  205 - 221 ( 200 - 224)
   INTEGRAL    Likelihood = -7.86   Transmembrane  276 - 292 ( 273 - 300)
   INTEGRAL    Likelihood = -6.32   Transmembrane  159 - 175 ( 154 - 176)
   INTEGRAL    Likelihood = -6.05   Transmembrane  236 - 252 ( 232 - 264)
   INTEGRAL    Likelihood = -5.95   Transmembrane   42 -  58 (  38 -  60)
   INTEGRAL    Likelihood = -5.84   Transmembrane  120 - 136 ( 119 - 138)
   INTEGRAL    Likelihood = -4.35   Transmembrane  255 - 271 ( 253 - 274)
   INTEGRAL    Likelihood = -1.59   Transmembrane   66 -  82 (  66 -  85)

----- Final Results -----
              bacterial membrane --- Certainty=0.4503(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD36213 GB:AE001771 branched chain amino acid ABC transporter,
           permease protein [Thermotoga maritima]
 Identities = 119/332 (35%), Positives = 191/332 (56%), Gaps = 33/332 (9%)

Query: 12  LAIVVLDYLLISVLISMGIFNLYHIQIIETIGINVILAVGLNLIVGCSGQFSLGHAGFMA 71
           L +V L ++ + + ++    + Y ++++  I I  I+AV LNLI G +G FSLGHAGF+
Sbjct: 16  LTVVFLIFMALLLYLADRYMDSYKLRVVRLIAIYGIMAVSLNLINGITGIFSLGHAGFIL 75

Query: 72  IGAYAVAIIGVKMP-----------------TYVGFLIAILVGTLVAGGIALGVGIPTLR 114
           IGAY  +++ +                     +  F  A + G ++A   A  +G P LR
Sbjct: 76  IGAYTASLLTLSPEQKAMSFIIEPIVPWLANAHTDFFTATVAGGVLAAVFAFLIGWPVLR 135

Query: 115 LKGDYLAIATLGVAEIIRILLVNGGDITNGAAGIMGIPPFTTWSLVYGVAVVSLILAMNF 174
           L GDYLAIA+LG AE+IRI+ +N   ITNG  G+ GIP ++     YG   V+++   +
Sbjct: 136 LSGDYLAIASLGFAEVIRIIALNAISITNGPLGLKGIPEYSNIWWCYGWLFVTVLFMASL 195

Query: 175 LRSPLGRNTIAIREDEIAAESMGVDTTKVKVIVFVFGAILASIAGSLQAGYVGTVMPKDF 234
           + S  GR   AIRED IAAE+MG++  K +++ FV GA  A ++GSL A ++ T+ P+
Sbjct: 196 VNSSYGRALKAIREDRIAAEAMGINVFKHQLLSFVIGAFFAGVSGSLYAHWLTTIDPRTT 255

Query: 235 SF--MMSVNVLIIVVLGGLGSMTGTVLAAILLGLLNMLLQD--------------YASVR 278
           +   M++  VLI++VLGGLGS++G+++ A L  +L   L+D                 +R
Sbjct: 256 TLGPMLTFYVLIMIVLGGLGSISGSLIGAALFAILFEWLRDLEEPFTFFGIHVPGIKGMR 315

Query: 279 MIIYALALILIMIFRPSGLLGTKELTLSHLFR 310
           +++ +   IL+MIF   G++G +ELT ++L+R
Sbjct: 316 ILVISAIFILVMIFWQRGIMGREELTWNNLYR 347
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 699

A DNA sequence (GBSx0743) was identified in *S.agalactiae* <SEQ ID 2153> which encodes the amino acid sequence <SEQ ID 2154>. This protein is predicted to be branched chain amino acid ABC transporter, ATP-binding protein (livG). Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2057(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD36214 GB:AE001771 branched chain amino acid ABC transporter,
          ATP-binding protein [Thermotoga maritima]
 Identities = 136/271 (50%), Positives = 189/271 (69%), Gaps = 21/271 (7%)

Query: 3   LLEVKNLSKHFGGLTAVGDVSMKLHKGELIGLIGPNGAGKTTLFNLLTGVYLPSKGTISI 62
            LL + +++  FGGL AV D + ++ +GEL+GLIGPNGAGKTT+FN++TG+Y P+KG I
Sbjct: 11  LLLLDHVTMQFGGLVAVDDFTNEIREGELVGLIGPNGAGKTTVFNVITGIYTPTKGRIVF 70

Query: 63  DGKILNGRKPAKIASLGLGRTFQNIRLFKNMTVLDNVLVGLSNHHLSHPIASFLRLPK-- 120
            +   + G +P +I  LG+ RTFQNIRLF +MTVL+NVLV   +H LS+P A  + +
Sbjct: 71  NDIDITGLRPYQITHLGIARTFQNIRLFSDMTVLENVLVA-QHHVLSNPDADRILVKHGK 129

Query: 121 ------------------YYHSEKALRKKALELLEIFGLKAYQDALAKNLPYGKQRRLEI 162
                              Y   EK + ++A +L++  GL+      A +LPYG+QR+LEI
Sbjct: 130 PRKGHGRFWFWRAVTKIGYLKKEKEMVERAKDLIKRVGLEKVMYEKASSLPYGEQRKLEI 189

Query: 163 VRALATEPKILFLDEPAAGMNPQETAELTQLISQIKSDFDITIMLIEHDMNLVMQVTERI 222
             RALATEPK++ LDEPAAGMNP+ET +L + I QI+ DF++T++LIEHDM +VM + ERI
Sbjct: 190 ARALATEPKLILLDEPAAGMNPKETEDLMEFIKQIRKDFNLTVLLIEHDMKVVMGICERI 249

Query: 223 YVLEYGRLIAHGTPEEIKNNKRVIEAYLGGE 253
             V++YGR+IA GTP+EI+N+ RVIEAYLG E
Sbjct: 250 IVMDYGRIIAEGTPKEIQNDPRVIEAYLGRE 280
```

There is also homology to SEQ ID 644.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 700

A DNA sequence (GBSx0744) was identified in *S.agalactiae* <SEQ ID 2155> which encodes the amino acid sequence <SEQ ID 2156>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2216(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB52068 GB:AL109732 putative branched chain amino acid
          transport ATP-binding protein [Streptomyces coelicolor
          A3(2)]
 Identities = 136/233 (58%), Positives = 181/233 (77%)
```

```
Query: 3   MLKVENLSIHYGVIQAVNDVSFEVNQGEVVTLIGANGAGKTSILRTISGLVRPSQGSISF 62
           +L+VE+L + YG I+AV  +SF+V+ GEVVTLIG NGAGKT+ LRT+SGL++P  G I F
Sbjct: 4   LLEVEDLRVAYGKIEAVKGISFKVDAGEVVTLIGTNGAGKTTTLRTLSGLLKPVGGQIRF 63

Query: 63  MGKPIHKLAARKIVGNGLAQVPEGRHVFSSLSVMENLEMGAFLQKDREQNQKMLKKVFDR 122
            GK + K+ A +IV  GLA  PEGRH+F  +++ +NL +GAFL+ DR   +K +++ +D
Sbjct: 64  GGKSLKKVPAHQIVSLGLAHSPEGRHIFPRMTIEDNLRLGAFLRSDRPGIEKDIQRAYDL 123

Query: 123 FPRLEERKNQDAATLSGGEQQMLAMGRALMSRPKLLLLDEPSMGLAPIFIQEIFNIIEDI 182
           FP L ER+ Q A TLSGGEQQMLAMGRALMS+PKLL+LDEPSMGL+PI +Q+I   I ++
Sbjct: 124 FPILGERRKQAAGTLSGGEQQMLAMGRALMSQPKLLMLDEPSMGLSPIMMQKIMATIAEL 183

Query: 183 KKQGTTVLLVEQNANKALTIADKAYVLETGKVVLSGTGKELLVSDQVRKAYLG 235
           K QGTT+LLVEQNA  AL++AD  +V+E G +VLSG+G++LL  + VRKAYLG
Sbjct: 184 KSQGTTILLVEQNAQAALSLADHGHVMEVGNIVLSGSGQDLLHDESVRKAYLG 236
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 701

A DNA sequence (GBSx0745) was identified in *S.agalactiae* <SEQ ID 2159> which encodes the amino acid sequence <SEQ ID 2160>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0415(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD36216 GB:AE001771 conserved hypothetical protein [Thermotoga maritima]
 Identities = 72/166 (43%), Positives = 116/166 (69%), Gaps = 2/166 (1%)

Query: 1   MPVKDFMTKKLVYVSPDTTVAEAADLLREHHLRRLPVVENDQLVGLVTEGTMAEAQPSKA 60
           M VKDFMT+  + ++P+T+ +EA  L++++ ++RL V++N+++VG+VTE  +  A PSKA
Sbjct: 1   MLVKDFMTRNPITIAPETSFSEALKLMKQNKIKRLIVMKNEKIVGIVTEKDLLYASPSKA 60

Query: 61  TSLSIYEMNYLLNKTKIRDIMIKDIVTVSQYASLEDAIYLMMSRKIGVLPVVDN-GQLYG 119
           T+L+I+E++YLL+K KI +IM KD+VTV++   +EDA  +M  + I  LPVVD+ G+L G
Sbjct: 61  TTLNIWELHYLLSKLKIEEIMTKDVVTVNENTPIEDAARIMEEKDISGLPVVDDAGRLVG 120

Query: 120 IVTDRDVFKAFLEIAGYGQE-SYRLVILADEGIGVLSKVLNRLSSA 164
           I+T  D+FK F+EI G  +E + R  +   +  G L +V  R+  A
Sbjct: 121 IITQTDIFKVFVEIFGTKREGTIRYTMEMPDKPGELLEVAKRIYEA 166
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 702

A DNA sequence (GBSx0746) was identified in *S.agalactiae* <SEQ ID 2163> which encodes the amino acid sequence <SEQ ID 2164>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5585(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 703

A DNA sequence (GBSx0747) was identified in *S.agalactiae* <SEQ ID 2165> which encodes the amino acid sequence <SEQ ID 2166>. This protein is predicted to be a transposase. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.65   Transmembrane   53 -  69 (  53 -  70)

----- Final Results -----
               bacterial membrane --- Certainty=0.1659(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA85003 GB:U28972 SpV1 ORF3; putative transposase [Spiroplasma citri]
 Identities = 49/154 (31%), Positives = 80/154 (51%), Gaps = 11/154 (7%)

Query: 39  WLEMDTVIGRIGGKVLLTFNVAFCNFIFAKLMDSKTAIETAKHIQ--VIKRTLYDNKRDF 96
           WLEMDTV+G+     +L          FA +++ TA E  K  +  +IK  L    +
Sbjct: 174 WLEMDTVVGKDHKSAILVLVEQLSKKYFAIKLENHTAREVEKKFKDIIIKNNLIGKIKG- 232

Query: 97  FELFPVILTDNGGEFARVDDIEIDVCGQSQLFFCDPNRSDQKARIEKNHTLVRDILPKGT 156
                 I+TD G EF++  ++EI    ++Q++FCD     QK  IE  ++ +R   PKGT
Sbjct: 233 ------IITDRGKEFSKWREMEI--FAETQVYFCDAGSPQQKPLIEYMNSELRHWFPKGT 284

Query: 157 SFDNLTQEDINLALSHINSVKRQALNGKTAYELF 190
            F+ ++Q+ I+  ++ IN   R  LN  ++ E+F
Sbjct: 285 DFNKVSQKQIDWVVNVINDKLRPCLNWISSKEMF 318
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 704

A DNA sequence (GBSx0748) was identified in *S.agalactiae* <SEQ ID 2167> which encodes the amino acid sequence <SEQ ID 2168>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3116(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10055> which encodes amino acid sequence <SEQ ID 10056> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 705

A DNA sequence (GBSx0749) was identified in *S.agalactiae* <SEQ ID 2169> which encodes the amino acid sequence <SEQ ID 2170>. This protein is predicted to be thymidylate kinase (tmk). Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1876(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10053> which encodes amino acid sequence <SEQ ID 10054> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB03761 GB:AP001507 thymidylate kinase [Bacillus halodurans]
 Identities = 112/210 (53%), Positives = 148/210 (70%), Gaps = 1/210 (0%)

Query: 17  MKKGLMISFEGPDGAGKTTVLEAVLPLLREKLSQDILTTREPGGVTISEEIRHIILDVKH 76
           M KG  I+ EG +GAGKT+ L+A+  +LRE     ++ TREPGG+ I+E+IR IILDV H
Sbjct: 1   MTKGCFITVEGGEGAGKTSALDAIEEMLREN-GLSVVRTREPGGIPIAEQIRSIILDVDH 59

Query: 77  TQMDKKTELLLYMAARRQHLVEKVLPALEEGKIVLMDRFIDSSVAYQGSGRGLDKSHIKW 136
           T+MD +TE LLY AARRQHLVEKVLPALE G +VL DRFIDSS+AYQG  RG+    I
Sbjct: 60  TRMDPRTEALLYAAARRQHLVEKVLPALEAGHVVLCDRFIDSSLAYQGYARGIGFEDILA 119

Query: 137 LNDYATDSHKPDLTLYFDVPSEVGLERIQKSVQREVNRLDLEQLDMHQRVRQGYLELADS 196
           +N++A +   PDLTL F V  +VGL RI +   RE NRLD E L  HQ+V++GY  + ++
Sbjct: 120 INEFAIEGRYPDLTLLFRVDPDVGLSRIHRDQSREQNRLDQEALTFHQKVKEGYERIVET 179

Query: 197 EPNRIVTIDASQQLDEVIAETFSIILDRIN 226
            P R+V IDA+Q  D+V+A+   +I  R++
Sbjct: 180 YPERVVEIDANQSFDQVVADAVRMIKQRLS 209
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2171> which encodes the amino acid sequence <SEQ ID 2172>. Analysis of this protein sequence reveals the following:

```
     Possible site: 56
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.75   Transmembrane  215 - 231 ( 215 - 231)

----- Final Results -----
               bacterial membrane --- Certainty=0.1298(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB03761 GB:AP001507 thymidylate kinase [Bacillus halodurans]
 Identities = 109/205 (53%), Positives = 148/205 (72%), Gaps = 1/205 (0%)

Query: 22  MITGKLITVEGPDGAGKTTVLEQLIPLLKQKVAQDILTTREPGGVAISEHIRELILDINH 81
```

```
           M  G  ITVEG +GAGKT+ L+ +  +L++     ++ TREPGG+ I+E IR +ILD++H
Sbjct: 1   MTKGCFITVEGGEGAGKTSALDAIEEMLREN-GLSVVRTREPGGIPIAEQIRSIILDVDH 59

Query: 82  TAMDPKTELLLYIAARRQHLVEKVLPALEAGQLVFIDRFIDSSVAYQGAGRGLIKADIQW 141
           T MDP+TE LLY AARRQHLVEKVLPALEAG +V  DRFIDSS+AYQG  RG+   DI
Sbjct: 60  TRMDPRTEALLYAAARRQHLVEKVLPALEAGHVVLCDRFIDSSLAYQGYARGIGFEDILA 119

Query: 142 LNEFATDGLEPDLTLYFDVPSEIGLARINANQQREVNRLDLETIEIHQRVRKGYLALAKE 201
           +NEFA +G  PDLTL F V  ++GL+RI+ +Q RE NRLD E +  HQ+V++GY  + +
Sbjct: 120 INEFAIEGRYPDLTLLFRVDPDVGLSRIHRDQSREQNRLDQEALTFHQKVKEGYERIVET 179

Query: 202 HPKRIVTIDATKPLKEVVSVALEHV 226
           +P+R+V IDA +   +VV+ A+  +
Sbjct: 180 YPERVVEIDANQSFDQVVADAVRMI 204
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 145/219 (66%), Positives = 181/219 (82%)

Query: 4   FDRIVVIINKGCTMKKGLMISFEGPDGAGKTTVLEAVLPLLREKLSQDILTTREPGGVTI 63
           FD+I ++ ++G  M  G +I+ EGPDGAGKTTVLE ++PLL++K++QDILTTREPGGV I
Sbjct: 9   FDKIELLKSEGNKMITGKLITVEGPDGAGKTTVLEQLIPLLKQKVAQDILTTREPGGVAI 68

Query: 64  SEEIRHIILDVKHTQMDKKTELLLYMAARRQHLVEKVLPALEEGKIVLMDRFIDSSVAYQ 123
           SE IR +ILD+ HT MD KTELLLY+AARRQHLVEKVLPALE G++V +DRFIDSSVAYQ
Sbjct: 69  SEHIRELILDINHTAMDPKTELLLYIAARRQHLVEKVLPALEAGQLVFIDRFIDSSVAYQ 128

Query: 124 GSGRGLDKSHIKWLNDYATDSHKPDLTLYFDVPSEVGLERIQKSVQREVNRLDLEQLDMH 183
           G+GRGL K+ I+WLN++ATD  +PDLTLYFDVPSE+GL RI  + QREVNRLDLE +++H
Sbjct: 129 GAGRGLIKADIQWLNEFATDGLEPDLTLYFDVPSEIGLARINANQQREVNRLDLETIEIH 188

Query: 184 QRVRQGYLELADSEPNRIVTIDASQQLDEVIAETFSIIL 222
           QRVR+GYL LA   P RIVTIDA++ L EV++     +L
Sbjct: 189 QRVRKGYLALAKEHPKRIVTIDATKPLKEVVSVALEHVL 227
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 706

A DNA sequence (GBSx0750) was identified in *S.agalactiae* <SEQ ID 2173> which encodes the amino acid sequence <SEQ ID 2174>. This protein is predicted to be DNA polymerase III delta' subunit (dnaZX). Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2603(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB03763 GB:AP001507 DNA polymerase III delta' subunit [Bacillus halodurans]
 Identities = 78/189 (41%), Positives = 113/189 (59%), Gaps = 3/189 (1%)

Query: 2   DLKRTQPKLLEKFNTILQSDRMSHAYLFSGNFAS--LDMALYLAQSQFCEKRQSGLPCQE 59
           +L + QP +       L   R++HAY+F GN  +    MAL+LA+S FC +R    PCQ
Sbjct: 5   NLAKNQPFVATMLKNSLAKGRLAHAYIFDGNRGTGKKRMALHLAKSFFCAQRAGVEPCQT 64

Query: 60  CRACRLIANGEFSDVKIIEPQGQLIKTETIKELTKDFSRSGFEGKSQVFIIKDCEKMHVN 119
           C+ C+ I +G   DV  IEP GQ IK   ++ L K+FS  G E   +V+I+   +KM  +
Sbjct: 65  CKECKRIEHGNHPDVHFIEPDGQSIKKHQVEHLQKEFSYRGMESAKKVYIVNHADKMTTS 124
```

```
Query: 120 AANSLLKFIEEPQSSSYVILLTNDENNVLPTIKSRTQIFRF-PKQLDMLVHQAEQAGLLK 178
           AANSLLKF+EEP + +  ILLT    N+LPTIKSR+Q+  F P ++        E+ G+ +
Sbjct: 125 AANSLLKFLEEPLADTVAILLTEQLQNMLPTIKSRSQVLSFAPLEVQAFAKLLEEEGISE 184

Query: 179 SQASLLAQV 187
           S ++LLA +
Sbjct: 185 SVSNLLASL 193
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2175> which encodes the amino acid sequence <SEQ ID 2176>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2685(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 151/290 (52%), Positives = 213/290 (73%), Gaps = 3/290 (1%)

Query: 1    MDLKRTQPKLLEKFNTILQSDRMSHAYLFSGNFASLDMALYLAQSQFCEKRQSGLPCQEC 60
            MDL +  P + + F TIL+ DR++HAYLFSG+FA+ +MAL+LA+  FCE+++   PC  C
Sbjct: 1    MDLAQKAPNVYQAFQTILKKDRLNHAYLFSGDFANEEMALFLAKVIFCEQKKDQTPCGHC 60

Query: 61   RACRLIANGEFSDVKIIEPQGQLIKTETIKELTKDFSRSGFEGKSQVFIIKDCEKMHVNA 120
            R+C+LI  G+F+DV ++EP GQ+IKT+ +KE+  +FS++G+E K QVFIIKDC+KMH+NA
Sbjct: 61   RSCQLIEQGDFADVTVLEPTGQVIKTDVVKEMMANFSQTGYENKRQVFIIKDCDKMHINA 120

Query: 121 ANSLLKFIEEPQSSSYVILLTNDENNVLPTIKSRTQIFRFPKQLDMLVHQAEQAGLLKSQ 180
            ANSLLK+IEEPQ  +Y+ LLTND+N VLPTIKSRTQ+F+FPK    L   A++ GLL  Q
Sbjct: 121 ANSLLKYIEEPQGEAYIFLLTNDDNKVLPTIKSRTQVFQFPKNEAYLYQLAQEKGLLNHQ 180

Query: 181 ASLLAQVADDPKHLEILLTNKKLLDYLNLSQQFVTTLAKDRQTAYLEVSRLTSQVVDKND 240
            A L+A++A +  HLE LL   KLL+ +  +++FV+   KD+  AYL ++RL     +K +
Sbjct: 181 AKLVAKLATNTSHLERLLQTSKLLELITQAERFVSIWLKDQLQAYLALNRLVQLATEKEE 240

Query: 241 QAFVFQWLTIMLAKE---GQLYDLENTYRAQQMWKSNVSFQNSLEYMVLS 287
            Q  V   LT++LA+E     L  LE  Y+A+ MW+SNV+FQN+LEYMV+S
Sbjct: 241 QDLVLTLLTLLLARERAQTPLTQLEAVYQARLMWQSNVNFQNTLEYMVMS 290
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 707

A DNA sequence (GBSx0751) was identified in *S.agalactiae* <SEQ ID 2177> which encodes the amino acid sequence <SEQ ID 2178>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2016(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB03765 GB:AP001507 unknown conserved protein in B. subtilis
           [Bacillus halodurans]
 Identities = 45/116 (38%), Positives = 62/116 (52%), Gaps = 8/116 (6%)
```

```
Query: 1   MDKKDLFDAFDDFSQNLLVGLSEIETMKKQIQKLLEENTVLRIENGKLRERLSVIEAET- 59
           M+KK +F       + +     E+  +K+Q+  L+EEN  L IEN  LRERL   E E
Sbjct: 1   MNKKAIFTQVSQLEERIGELHRELGGLKEQLAYLIEENHFLTIENEHLRERLGEPELEET 60

Query: 60  ---ETAVKNSK----QGRELLEGIYNDGFHICNTFYGQRRENDEECAFCIELLYRD 108
              E   K  K    +G + L  +Y +GFHICNT YG  R+N E+C FC+  L +D
Sbjct: 61  EEKEQVTKERKPFVGEGYDNLARLYQEGFHICNTHYGSLRKNGEDCLFCLSFLNQD 116
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2179> which encodes the amino acid sequence <SEQ ID 2180>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0700(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 75/107 (70%), Positives = 89/107 (83%), Gaps = 1/107 (0%)

Query: 1   MDKKDLFDAFDDFSQNLLVGLSEIETMKKQIQKLLEENTVLRIENGKLRERLSVIEAETE 60
           ++KK+LFDAFD FSQNL+V L+EIE MKKQ+Q L+EENT+LR+EN KLRERLS +E ET
Sbjct: 1   VNKKELFDAFDGFSQNLMVTLAEIEAMKKQVQSLVEENTILRLENTKLRERLSHLEHET- 59

Query: 61  TAVKNSKQGRELLEGIYNDGFHICNTFYGQRRENDEECAFCIELLYR 107
            A   SKQ ++ LEGIY++GFHICN FYGQRRENDEEC FC ELL R
Sbjct: 60  VAKNPSKQRKDHLEGIYDEGFHICNFFYGQRRENDEECMFCRELLDR 106
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 708

A DNA sequence (GBSx0752) was identified in *S.agalactiae* <SEQ ID 2181> which encodes the amino acid sequence <SEQ ID 2182>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.28   Transmembrane  119 - 135 ( 119 - 135)

----- Final Results -----
               bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10051> which encodes amino acid sequence <SEQ ID 10052> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB03768 GB:AP001507 unknown conserved protein [Bacillus halodurans]
 Identities = 138/287 (48%), Positives = 189/287 (65%), Gaps = 2/287 (0%)

Query: 4   MQVQKSFKSNIHYGTLYLVPTPIGNLDDMTFRAIRILREVDFICAEDTRNTGLLLKHFDI 63
           M+ Q+S++     GTLYLV TPIGNL+D+TFRAIR L+E D I AEDTR T  LL HFDI
Sbjct: 1   MKTQQSYQQRDDKGTLYLVATPIGNLEDVTFRAIRTLKEADQIAAEDTRQTKKLLNHFDI 60

Query: 64  TTKQISFHEHNAYDKISGLIDLLKEGKSLAQVSDAGMPSISDPGHDLVKAAIEGDIPVVS 123
            TK +S+HEHN       LID L EG+++A VSDAGMP+ISDPG++LV +AI+  I V+
Sbjct: 61  ATKLVSYHEHNKETMGKRLIDDLIEGRTIALVSDAGMPAISDPGYELVVSAIKEGIAVIP 120
```

```
Query: 124 IPGASAGITALIASGLAPQPHIFYGFLPRKKGQQITFFETKQDYPETQIFYESPFRVSDT 183
           IPGA+A +TALIASGL  +   F GFLPR+K Q+    E  +    T IFYESP R+ DT
Sbjct: 121 IPGANAAVTALIASGLPTESFQFIGFLPRQKKQRRQALEETKPTKATLIFYESPHRLKDT 180

Query: 184 LKHMKEIYGDRQVVLVRELTKLYEEYQRGTISQLLEHIEKVPLKGECLIIVDGKRDTERV 243
           L  M  I G+R V + RELTK YEE+ RGT+ + +    +  +KGE  +IV+G  +
Sbjct: 181 LDDMLLILGNRHVSICRELTKTYEEFLRGTLEEAVHWAREATIKGEFCLIVEGNGEKVEP 240

Query: 244 KDS--SQQDPLVLVKEYIANGDKTNQAIKKVAKEFNLNRQELYASFH 288
           ++       P+  V+ YIA G ++ +AIK+VA +  + ++++Y  +H
Sbjct: 241 EEVWWESLSPVQHVEHYIALGFRSKEAIKQVATDRGVPKRDIYNIYH 287
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2183> which encodes the amino acid sequence <SEQ ID 2184>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.09   Transmembrane  116 - 132 ( 116 - 134)

----- Final Results -----
              bacterial membrane --- Certainty=0.2635(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB03768 GB:AP001507 unknown conserved protein [Bacillus halodurans]
 Identities = 139/287 (48%), Positives = 189/287 (65%), Gaps = 2/287 (0%)

Query: 1   MQVQKSFKDKKTSGTLYLVPTPIGNLQDMTFRAVATLKEVDFICAEDTRNTGLLLKHFDI 60
           M+ Q+S++ +   GTLYLV TPIGNL+D+TFRA+ TLKE D I AEDTR T  LL HFDI
Sbjct: 1   MKTQQSYQQRDDKGTLYLVATPIGNLEDVTFRAIRTLKEADQIAAEDTRQTKKLLNHFDI 60

Query: 61  ATKQISFHEHNAYEKIPDLIDLLISGRSLAQVSDAGMPSISDPGHDLVKAAIDSDIAVVA 120
           ATK +S+HEHN       LID LI GR++A VSDAGMP+ISDPG++LV +AI   IAV+
Sbjct: 61  ATKLVSYHEHNKETMGKRLIDDLIEGRTIALVSDAGMPAISDPGYELVVSAIKEGIAVIP 120

Query: 121 LPGASAGITALIASGLAPQPHVFYGFLPRKAGQQKAFFEDKHHYPETQMFYESPYRIKDT 180
           +PGA+A +TALIASGL  +   F GFLPR+  Q++   E+      T +FYESP+R+KDT
Sbjct: 121 IPGANAAVTALIASGLPTESFQFIGFLPRQKKQRRQALEETKPTKATLIFYESPHRLKDT 180

Query: 181 LTNMLACYGDRQVVLVRELTKLFEEYQRGSISEILSYLEETPLKGECLLIVA--GAQADS 238
           L +ML   G+R V + RELTK +EE+ RG++ E + +  E  +KGE  LIV   G + +
Sbjct: 181 LDDMLLILGNRHVSICRELTKTYEEFLRGTLEEAVHWAREATIKGEFCLIVEGNGEKVEP 240

Query: 239 EVELTADVDLVSLVQKEIQAGAKPNQAIKTIAKAYQVNRQELYQQFH 285
           E      +  V  V+  I  G +  +AIK +A    V ++++Y  +H
Sbjct: 241 EEVWWESLSPVQHVEHYIALGFRSKEAIKQVATDRGVPKRDIYNIYH 287
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 208/287 (72%), Positives = 238/287 (82%)

Query: 4   MQVQKSFKSNIHYGTLYLVPTPIGNLDDMTFRAIRILREVDFICAEDTRNTGLLLKHFDI 63
           MQVQKSFK     GTLYLVPTPIGNL DMTFRA+  L+EVDFICAEDTRNTGLLLKHFDI
Sbjct: 1   MQVQKSFKDKKTSGTLYLVPTPIGNLQDMTFRAVATLKEVDFICAEDTRNTGLLLKHFDI 60

Query: 64  TTKQISFHEHNAYDKISGLIDLLKEGKSLAQVSDAGMPSISDPGHDLVKAAIEGDIPVVS 123
            TKQISFHEHNAY+KI  LIDLL  G+SLAQVSDAGMPSISDPGHDLVKAAI+ DI VV+
Sbjct: 61  ATKQISFHEHNAYEKIPDLIDLLISGRSLAQVSDAGMPSISDPGHDLVKAAIDSDIAVVA 120

Query: 124 IPGASAGITALIASGLAPQPHIFYGFLPRKKGQQITFFETKQDYPETQIFYESPFRVSDT 183
           +PGASAGITALIASGLAPQPH+FYGFLPRK GQQ  FFE K  YPETQ+FYESP+R+ DT
Sbjct: 121 LPGASAGITALIASGLAPQPHVFYGFLPRKAGQQKAFFEDKHHYPETQMFYESPYRIKDT 180

Query: 184 LKHMKEIYGDRQVVLVRELTKLYEEYQRGTISQLLEHIEKVPLKGECLIIVDGKRDTERV 243
           L +M   YGDRQVVLVRELTKL+EEYQRG+IS++L ++E+ PLKGECL+IV G +    V
```

```
Sbjct: 181 LTNMLACYGDRQVVLVRELTKLFEEYQRGSISEILSYLEETPLKGECLLIVAGAQADSEV 240

Query: 244 KDSSQQDPLVLVKEYIANGDKTNQAIKKVAKEFNLNRQELYASFHDL 290
           + ++  D + LV++ I  G K NQAIK +AK + +NRQELY  FHDL
Sbjct: 241 ELTADVDLVSLVQKEIQAGAKPNQAIKTIAKAYQVNRQELYQQFHDL 287
```

A related GBS gene <SEQ ID 8643> and protein <SEQ ID 8644> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:     -6.92
GvH: Signal Score (-7.5): -9.26
     Possible site: 48
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -1.28 threshold:  0.0
   INTEGRAL    Likelihood = -1.28   Transmembrane  118 - 134 ( 118 - 134)
   PERIPHERAL  Likelihood =  6.89      32
 modified ALOM score:   0.76

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00263(310 - 1164 of 1470)
EGAD|17863|BS0036(2 - 289 of 292) hypothetical 33.0 kd protein in xpac-abrb intergenic
region {Bacillus subtilis} OMNI|NT01BS0044 conserved hypothetical protein
SP|P37544|YABC_BACSU HYPOTHETICAL 33.0 KDA PROTEIN IN XPAC-ABRB INTERGENIC REGION.
GP|467425|dbj|BAA05271.1||D26185 unknown {Bacillus subtilis}
GP|2632303|emb|CAB11812.1||Z99104 similar to hypothetical proteins {Bacillus subtilis}
PIR|S66065|S66065 conserved hypothetical protein yabC - Bacillus subtilis
%Match = 24.5
%Identity = 45.8  %Similarity = 65.7
Matches = 131  Mismatches = 97  Conservative Sub.s = 57

123       153       183       213       243       273       303       333
CSTH*KW*TS*ASERY*SRNRNCS*KF*TRKRITRRHLQ*WLSHL*YFLWSTS*K*RRMCFLY*III*RLMEMQVQKSFK
                                                                        :: | ||
                                                                       MLRRQMSFN

363       393       423       453       483       513       543       573
SNIHYGTLYLVPTPIGNLDDMTFRAIRILREVDFICAEDTRNTGLLLKHFDITTKQISFHEHNAYDKISGLIDLLKEGKS
     | |||||||||||:|||||||  |: || | ||||| |  |   ::| |  :|:||||       :|: || ||:
GKSDMGILYLVPTPIGNLEDMTFRAIDTLKSVDAIAAEDTRQTKKLCHVYEIETPLVSYHEHNKESSGHKIIEWLKSGKN
         20        30        40        50        60        70        80

603       633       663       693       723       753       783       813
LAQVSDAGMPSISDPGHDLVKAAIEGDIPVVSIPGASAGITALIASGLAPQPHIFYGFLPRKKGQQITFFETKQDYPETQ
:| |||||:|:||||| ::||   :    || :|||:| :|||||||: |||  ||||| |:| ::   :|  :   ||
IALVSDAGLPTISDPGAEIVKDFTDIGGYVVPLPGANAALTALIASGIVPQPFFFYGFLNRQKKEKKKELEALKKRQETI
        100       110       120       130       140       150       160

843       873       903       933       963       993      1023      1053
IFYESPFRVSDTLKHMKEIYGDRQVVLVRELTKLYEEYQRGTISQLLEHIEKVPLKGECLIIVDGKRDTERVKDSSQQDP
||||:| |: :||  | || |||:: : ||||| |||: |||||:::    :  ::||  ::|:|  : |  ::    :
IFYEAPHRLKETLSAMAEILGDREIAVTRELTKKYEEFIRGTISEVIGWANEDQIRGEFCLVVEGSNNEEVDEEEQWWET
        180       190       200       210       220       230       240

1074      1104      1134      1164      1194      1224      1254      1284
LVL---VKEYIANGDKTNQAIKKVAKEFNLNRQELYASFHDL*VII*KGCQRKIWQPFIISDLAIGIKK*DTSNFLKIFN
|     |: ||: |  : :|||| | : |: ::|:| ::|
LTAKEHVEHYISKGATSKEAIKKAAVDRNVPKREVYDAYHIKQ
        260       270       280       290
```

SEQ ID 8644 (GBS343) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 11; MW 35.4kDa).

The GBS343-His fusion product was purified (Figure 215, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 277), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 709

A DNA sequence (GBSx0753) was identified in *S.agalactiae* <SEQ ID 2185> which encodes the amino acid sequence <SEQ ID 2186>. This protein is predicted to be bA483F11.3 (cutC). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2568(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB88199 GB:AL133353 bA483F11.3 (CGI-32 protein ) [Homo sapiens]
 Identities = 79/203 (38%), Positives = 116/203 (56%), Gaps = 7/203 (3%)

Query: 3   LREFCAENLTDLTRLDKAIISRVELCDNLAVGGTTPSYGVIKEANQYLHEKGISVAVMIR 62
           L E C +++       ++    R+ELC  L+ GGTTPS GV++   Q +    I V VMIR
Sbjct: 27  LMEVCVDSVESAVNAERGGADRIELCSGLSEGGTTPSMGVLQVVKQSVQ---IPVFVMIR 83

Query: 63  PRGGNFVYNDLELRIMEEDILRAVELESDALVLGILTSNNHIDTEAIEQLLPATQGLPLV 122
           PRGG+F+Y+D E+ +M+ DI  A    +D LV G LT + HID E    L+   + LP+
Sbjct: 84  PRGGDFLYSDREIEVMKADIRLAKLYGADGLVFGALTEDGHIDKELCMSLMAICRPLPVT 143

Query: 123 FHMAFDVIPKSDQKKSIDQLVALGFTRILLHGSSNGEPIIENIKHIKALVEYANNRIEIM 182
           FH AFD++   D   +++ L+ LGF R+L  G  +    +E +  IK L+E A  RI +M
Sbjct: 144 FHRAFDMV--HDPMAALETLLTLGFERVLTSGCDSS--ALEGLPLIKRLIEQAKGRIVVM 199

Query: 183 VGGGVTAENYQYICQETGVKQAH 205
            GGG+T  N Q I + +G  + H
Sbjct: 200 PGGGITDRNLQRILEGSGATEFH 222
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2187> which encodes the amino acid sequence <SEQ ID 2188>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2372(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 143/208 (68%), Positives = 168/208 (80%)

Query: 2   ILREFCAENLTDLTRLDKAIISRVELCDNLAVGGTTPSYGVIKEANQYLHEKGISVAVMI 61
           +++EFCAENLT L  LD   ISRVELCDNLAVGGTTPSYGVIKEA Q LH+K ISVA MI
```

```
Sbjct: 1   MIKEFCAENLTLLPTLDAGQISRVELCDNLAVGGTTPSYGVIKEACQLLHDKKISVATMI 60

Query: 62  RPRGGNFVYNDLELRIMEEDILRAVELESDALVLGILTSNNHIDTEAIEQLLPATQGLPL 121
            RPRGG+FVYNDLEL+ MEEDIL+AVE  SDALVLG+LT+ N +DT+AIEQLLPATQGLPL
Sbjct: 61  RPRGGDFVYNDLELKAMEEDILKAVEAGSDALVLGLLTTENQLDTDAIEQLLPATQGLPL 120

Query: 122 VFHMAFDVIPKSDQKKSIDQLVALGFTRILLHGSSNGEPIIENIKHIKALVEYANNRIEI 181
            VFHMAFD IP   Q +++DQL+  GF R+L HGS    PI +N++ +K+LV YAN RIEI
Sbjct: 121 VFHMAFDRIPTDHQHQALDQLIDYGFVRVLTHGSPEATPITDNVEQLKSLVTYANKRIEI 180

Query: 182 MVGGGVTAENYQYICQETGVKQAHGTRI 209
            M+GGG+TAEN Q + Q TG    HGT+I
Sbjct: 181 MIGGGITAENCQSLSQLTGTAIVHGTKI 208
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 710

A DNA sequence (GBSx0754) was identified in *S.agalactiae* <SEQ ID 2189> which encodes the amino acid sequence <SEQ ID 2190>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1216(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA12206 GB:D84061 phosphoserine aminotransferase [Spinacia
           oleracea]
 Identities = 65/109 (59%), Positives = 79/109 (71%), Gaps = 1/109 (0%)

Query: 3   IYNFSAGPAVLPKPVLVKAQSELLNYQGSSMSVLEVSHRSKEFDDIIKGAERYLRDLMGI 62
            ++NF+AGPAVLP+ VL KAQSELLN++GS MSV+E+SHR KEF  II  AE  LR L+ I
Sbjct: 69  VFNFAAGPAVLPENVLQKAQSELLNWRGSGMSVMEMSHRGKEFTSIIDKAEADLRTLLNI 128

Query: 63  PDNYKVIFLQGGASLQFSMIPLNIARGRKAY-YHVAGSWGEKSLYRGCK 110
            P +Y V+FLQGGAS QFS IPLN+     A  Y V GSWG+K+     K
Sbjct: 129 PSDYTVLFLQGGASTQFSAIPLNLCTPDSAVDYIVTGSWGDKAAKEAAK 177
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 711

A DNA sequence (GBSx0755) was identified in *S.agalactiae* <SEQ ID 2191> which encodes the amino acid sequence <SEQ ID 2192>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 712

A DNA sequence (GBSx0756) was identified in *S.agalactiae* <SEQ ID 2193> which encodes the amino acid sequence <SEQ ID 2194>. This protein is predicted to be phosphoserine aminotransferase (serC). Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3380(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10049> which encodes amino acid sequence <SEQ ID 10050> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF94318 GB:AE004196 phosphoserine aminotransferase [Vibrio cholerae]
 Identities = 104/210 (49%), Positives = 152/210 (71%), Gaps = 3/210 (1%)

Query: 4   NNTIEGTSLYDIPKTNEVPVIADMSSNILAVKYKVEDFAMIYAGAQKNIGPAGVTVVIIR 63
           N TI+G  + D+P T++ P++ADMSS IL+ +  V  + +IYAGAQKNIGPAG+ + I+R
Sbjct: 170 NETIDGIEINDLPVTDK-PIVADMSSTILSREIDVSKYGVIYAGAQKNIGPAGICIAIVR 228

Query: 64  EDMIN-EEPTLSSMLDYKIQSDAGSLYNTPPAYSIYIAKLVFEWVKSLGGVDAMEKANRE 122
           +D+++     L  +L+YKI ++  S++NTPP ++ Y++ LVF+W+K+ GGV A+E+ NR
Sbjct: 229 DDLLDLASDLLPGVLNYKILAEQESMFNTPPTFAWYLSGLVFQWLKAQGGVKAIEEVNRA 288

Query: 123 KSGLLYDYIDSSEFYSNPVRDKKSRSLCNIPFITINKDLDEKFVKEATERGFKNIKGHRS 182
           K+ LLY YIDSS+FY N +    +RSL N+PF     +LD+ F++ A  RG  ++KGHR
Sbjct: 289 KAALLYGYIDSSDFYRNEIH-PDNRSLMNVPFQLAKPELDDTFLELAEARGLVSLKGHRV 347

Query: 183 VGGMRASLYNAFPKQGVIELIDFMKTFEAE 212
           VGGMRAS+YNA P +GV  L+DFMK FEA+
Sbjct: 348 VGGMRASIYNAMPLEGVQALVDFMKEFEAQ 377
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 713

A DNA sequence (GBSx0757) was identified in *S.agalactiae* <SEQ ID 2195> which encodes the amino acid sequence <SEQ ID 2196>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0466(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10047> which encodes amino acid sequence <SEQ ID 10048> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB73701 GB:AL139079 putative acetyltransferase [Campylobacter
          jejuni]
 Identities = 46/170 (27%), Positives = 78/170 (45%), Gaps = 13/170 (7%)

Query: 7   IRLAFPNEIDQIMLLIEEARAEIAKTGSDQWQKEDGYPNRNDIIDDILNGYAWVGIEDGM 66
           I+ A  +++  I+ + ++A   +      QW  ++ YPN     +DI     +V  E+
Sbjct: 6   IQKAVNKDLNSILEITKDALNAMKTMNFHQW--DENYPNEIVFQEDIQAQELYVFKENDE 63

Query: 67  LATYAAVIDGHE-EVYDAIYEGKWLHDNHRYLTFHRIAISNQFRGRGLAQTFLQGL---- 121
           +  +  + +  + E Y  +   K   D   YL  HR+A+    +G+G+AQ  L
Sbjct: 64  ILGFICINEKFKPEFYKQVIFNKNYDDKAFYL--HRLAVKQNAKGKGVAQKLLNFCENFA 121

Query: 122 IEGHKGPDFRCDTHEKNVTMQHILNKLGYQYCGKVPLDGVR---LAYQKI 168
           +E HK    R DTH KN  M  +  KL + +CG   +   +   LAY+KI
Sbjct: 122 LENHKA-SLRADTHSKNFPMNSLFKKLDFNFCGNFDIPNYQDPFLAYEKI 170
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 714

A DNA sequence (GBSx0758) was identified in *S.agalactiae* <SEQ ID 2197> which encodes the amino acid sequence <SEQ ID 2198>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2968(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 715

A DNA sequence (GBSx0759) was identified in *S.agalactiae* <SEQ ID 2199> which encodes the amino acid sequence <SEQ ID 2200>. This protein is predicted to be D-3-phosphoglycerate dehydrogenase (serA). Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3102(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10045> which encodes amino acid sequence <SEQ ID 10046> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB99020 GB:U67544 phosphoglycerate dehydrogenase (serA)
           [Methanococcus jannaschii]
 Identities = 102/313 (32%), Positives = 168/313 (53%), Gaps = 21/313 (6%)

Query: 31  ENPDAYIIRSQNLHNQDF---PSNLKAIARAGAGTNNIPIEEASAQGIVVFNTPGANANA 87
            ++ D  ++RS     +D       LK I RAG G +NI +E A+ +GI+V N P A++ +
Sbjct: 40  KDADVLVVRSGTKVTRDVIEKAEKLKVIGRAGVGVDNIDVEAATEKGIIVVNAPDASSIS 99

Query: 88  VKEAVIAALLLSARDYLGANRWVNTLTGTDIPKQIEAGKKAFAGNEIAGKKLGVIGLGAI 147
            V E  +  +L +AR         N    T   K+ E  +K F G E+ GK LGVIGLG I
Sbjct: 100 VAELTMGLMLAAAR---------NIPQATASLKRGEWDRKRFKGIELYGKTLGVIGLGRI 150

Query: 148 GARIANDARRLGMTVLGYDPYVSIETAWNISSHVQRVKEIKDIFETCDYITIHVPLTNET 207
            G ++   A+  GM ++GYDPY+  E A ++   V+ V +I ++ +  D+IT+HVPLT +T
Sbjct: 151 GQQVVKRAKAFGMNIIGYDPYIPKEVAESMG--VELVDDINELCKRADFITLHVPLTPKT 208

Query: 208 KHTFDAKAFSIMKKGTTIINFARAELVNNQELFEAIETGVVKRYITDFGDKE------LL 261
            +H    +  ++MKK   I+N AR  L++ + L+EA++ G ++    D  ++E      LL
Sbjct: 209 RHIIGREQIALMKKNAIIVNCARGGLIDEKALYEALKEGKIRAAALDVFEEEPPKDNPLL 268

Query: 262 NQKGITVFPHVGGSTDEAELNCAIMASQTIRCFMETGEITNSVNFPNVHQIQTAPFR-IT 320
                +   PH G ST+EA+     + ++ I+  +      N VN PN+ Q +    +
Sbjct: 269 TLDNVIGTPHQGASTEEAQKAAGTIVAEQIKKVLRGELAENVVNMPNIPQEKLGKLKPYM 328

Query: 321 LINKNVPNIVAKI 333
            L+ + + NIV ++
Sbjct: 329 LLAEMLGNIVMQV 341
```

There is also homology to SEQ ID 124.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 716

A DNA sequence (GBSx0760) was identified in *S.agalactiae* <SEQ ID 2201> which encodes the amino acid sequence <SEQ ID 2202>. This protein is predicted to be methylated-DNA--protein-cysteine S-methyltransferase (ogt). Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2460(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF96913 GB:AE004427 methylated-DNA--protein-cysteine
           S-methyltransferase [Vibrio cholerae]
 Identities = 73/156 (46%), Positives = 99/156 (62%), Gaps = 9/156 (5%)

Query: 7   YQSPLGEIRLLADNLGLSGLYFVGQKYDMLAVNQEEIVNMSNSYTLLGK--KWLDAYFSQ 64
            Y SPLG + L A + GL G++F  Q         E + +      +L K  + LD YFS
Sbjct: 7   YSSPLGPMTLQASSQGLLGVWFATQ-----TTQPEHLGDYVKECPILNKTIRQLDEYFSG 61

Query: 65  QNLP-SIPLSLRGTAFQTRVWQELQKIPFGDTKTYGELAKEL-NCQSAQAVGGAIGKNSI 122
            Q     +PL+  GTAFQ  VW  L KIP+G+  +Y +LA+ + N ++ +AVG A GKN I
```

```
Sbjct: 62  QRTQFELPLAASGTAFQQSVWHALCKIPYGEIWSYQQLAEAIGNPKAVRAVGLANGKNPI 121

Query: 123 SLIIPCHRVLGRYGQLTGYAGGLERKSWLLEYEKEK 158
           S+I+PCHRV+G+ GQLTGYAGGLERK++LLE EK +
Sbjct: 122 SIIVPCHRVVGKNGQLTGYAGGLERKAFLLELEKRR 157
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 717

A DNA sequence (GBSx0761) was identified in *S.agalactiae* <SEQ ID 2203> which encodes the amino acid sequence <SEQ ID 2204>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3137(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB07204 GB:AP001518 arsenate reductase [Bacillus halodurans]
 Identities = 56/107 (52%), Positives = 74/107 (68%), Gaps = 1/107 (0%)

Query: 3   TFYEYPKCTTCRSAKKELTELGLTFEAIDIKSNPPKVSLLKELLENSPYDLKKFFNTSGN 62
            TFY+YPKC TC+ AKK L + G+   ++ I   PP    LK+L E S  +LKKFFNTSG
Sbjct: 4   TFYQYPKCGTCQKAKKWLDQHGIEVNSVHIVEQPPSKEELKQLYEQSGLELKKFFNTSGK 63

Query: 63  SYRELGLKDKFDDLTLDQALDLLASDGMLIKRPLLVKDNKILQIGYR 109
             YRELGLKDK  + + D+ L+ LASDGMLIKRP+L   +K+  +G++
Sbjct: 64  KYRELGLKDKVKEASEDELLETLASDGMLIKRPILTDGDKV-TVGFK 109
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2205> which encodes the amino acid sequence <SEQ ID 2206>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3969(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 64/99 (64%), Positives = 79/99 (79%)

Query: 19  ELTELGLTFEAIDIKSNPPKVSLLKELLENSPYDLKKFFNTSGNSYRELGLKDKFDDLTL 78
            EL +L   FEAIDIK+NPPK   LK  +E S Y +K FFNTSGNSYRELGLKDK D L+L
Sbjct: 3   ELKQLVSDFEAIDIKANPPKAQDLKHWMETSGYTIKNFFNTSGNSYRELGLKDKIDQLSL 62

Query: 79  DQALDLLASDGMLIKRPLLVKDNKILQIGYRTKYKDLNL 117
            D+A +LLA+DGMLIKRP+L+KD  +LQ+GYR  Y++L+L
Sbjct: 63  DKAAELLATDGMLIKRPILIKDGNVLQVGYRKPYQELDL 101
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 718

A DNA sequence (GBSx0762) was identified in *S.agalactiae* <SEQ ID 2207> which encodes the amino acid sequence <SEQ ID 2208>. This protein is predicted to be exodeoxyribonuclease (exoA). Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1859(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA26879 GB:J04234 exodeoxyribonuclease [Streptococcus pneumoniae]
 Identities = 217/275 (78%), Positives = 245/275 (88%)

Query: 1    MKLISWNIDSLNAALTSESTRALMSRQVIDTLVAEDADIIAIQETKLSAKGPTKKHLEVL 60
            MKLISWNIDSLNAALTS+S RA +S++V+ TLVAE+ADIIAIQETKLSAKGPTKKH+E+L
Sbjct: 1    MKLISWNIDSLNAALTSDSARAKLSQEVLQTLVAENADIIAIQETKLSAKGPTKKHVEIL 60

Query: 61   ETYFPEYDLVWRSSVEPARKGYAGTMFLYRKGLNPIVSFPEIDAPTTMDNEGRIITLELE 120
            E  FP Y+  WRSS EPARKGYAGTMFLY+K L P +SFPEI AP+TMD EGRIITLE +
Sbjct: 61   EELFPGYENTWRSSQEPARKGYAGTMFLYKKELTPTISFPEIGAPSTMDLEGRIITLEFD 120

Query: 121 NCYITQVYTPNAGDGLKRLADRQIWDIKYAEYLATLDSQKPVLATGDYNVAHKEIDLANP 180
              ++TQVYTPNAGDGLKRL +RQ+WD KYAEYLA LD +KPVLATGDYNVAH EIDLANP
Sbjct: 121 AFFVTQVYTPNAGDGLKRLEERQVWDAKYAEYLAELDKEKPVLATGDYNVAHNEIDLANP 180

Query: 181 SSNRRSAGFTAEERQGFTNLLAKGFTDTFRYLHGDVPNVYSWWAQRSRTSKINNTGWRID 240
            +SNRRS GFT EER GFTNLLA GFTDTFR++HGDVP  Y+WWAQRS+TSKINNTGWRID
Sbjct: 181 ASNRRSPGFTDEERAGFTNLLATGFTDTFRHVHGDVPERYTWWAQRSKTSKINNTGWRID 240

Query: 241 YWLTSNRVADKITKSEMIHSGDRQDHTPIILEIEL 275
            YWLTSNR+ADK+TKS+MI SG RQDHTPI+LEI+L
Sbjct: 241 YWLTSNRIADKVTKSDMIDSGARQDHTPIVLEIDL 275
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2209> which encodes the amino acid sequence <SEQ ID 2210>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2181(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 221/275 (80%), Positives = 251/275 (90%)

Query: 1    MKLISWNIDSLNAALTSESTRALMSRQVIDTLVAEDADIIAIQETKLSAKGPTKKHLEVL 60
            MKLISWNIDSLNAALT ES RAL+SR V+DTLVA+DADIIAIQETKLSAKGPTKKH+E L
Sbjct: 1    MKLISWNIDSLNAALTGESPRALLSRAVLDTLVAQDADIIAIQETKLSAKGPTKKHIETL 60

Query: 61   ETYFPEYDLVWRSSVEPARKGYAGTMFLYRKGLNPIVSFPEIDAPTTMDNEGRIITLELE 120
             +YFP Y  VWRSSVEPARKGYAGTMFLY+  LNP+++FPEI APTTMD EGRIITLE E
Sbjct: 61   LSYFPNYLHVWRSSVEPARKGYAGTMFLYKNTLNPVITFPEIGAPTTMDAEGRIITLEFE 120

Query: 121 NCYITQVYTPNAGDGLKRLADRQIWDIKYAEYLATLDSQKPVLATGDYNVAHKEIDLANP 180
            + ++TQVYTPNAGDGL+RL DRQIWD KYA+YL  LD+QKPVLATGDYNVAHKEIDLANP
Sbjct: 121 DFFVTQVYTPNAGDGLRRLDDRQIWDHKYADYLTELDAQKPVLATGDYNVAHKEIDLANP 180
```

```
Query: 181 SSNRRSAGFTAEERQGFTNLLAKGFTDTFRYLHGDVPNVYSWWAQRSRTSKINNTGWRID 240
           +SNRRS GFT EERQGFTNLLA+GFTDTFR++HGD+P+VY+WWAQRS+TSKINNTGWRID
Sbjct: 181 NSNRRSPGFTDEERQGFTNLLARGFTDTFRHVHGDIPHVYTWWAQRSKTSKINNTGWRID 240

Query: 241 YWLTSNRVADKITKSEMIHSGDRQDHTPIILEIEL 275
           YWL SNR+ DK+ +SEMI SG+RQDHTPI+L+I+L
Sbjct: 241 YWLASNRLVDKVKRSEMISSGERQDHTPILLDIDL 275
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 719

A DNA sequence (GBSx0763) was identified in *S.agalactiae* <SEQ ID 2211> which encodes the amino acid sequence <SEQ ID 2212>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.96   Transmembrane   28 -  44 (  22 -  49)

----- Final Results -----
               bacterial membrane --- Certainty=0.4185(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8645> which encodes amino acid sequence <SEQ ID 8646> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 5
McG: Discrim Score:     17.78
GvH: Signal Score (-7.5): -4.56
     Possible site: 55
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -7.96 threshold:  0.0
   INTEGRAL    Likelihood = -7.96   Transmembrane    8 -  24 (   2 -  29)
   PERIPHERAL  Likelihood =  9.28     138
 modified ALOM score:   2.09

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4185(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD11512 GB:U60828 unknown [Lactococcus lactis]
 Identities = 53/240 (22%), Positives = 102/240 (42%), Gaps = 24/240 (10%)

Query: 65  PTILIPGSSATQERFNSMLAQL----NQMGEKHSVLKLTVKKDNSIIYNGQISGNDHKPY 120
           PTI I GS       + ++ +L    N   +K  V+   + K+  +   GQIS ++  P
Sbjct: 64  PTIYIGGSGGNVTSIDWLVERLLPIKNISSQKSLVMTSNITKNYELKVEGQISQDNKYPI 123

Query: 121 IVIGFENNEDGYSNIKKQTKWLQIAMNDLQKKYKFKRFNAIGHSNGGLSWTIFLEDYYDS 180
           I         G ++ +  +K LQ  +  L + Y+    N +G+S+G      ++ D  ++
Sbjct: 124 IEFA---TVKGTNSGELFSKGLQKIIVYLTENYQVPWINLVGYSSGATGAVYYMMDTGNN 180

Query: 181 DEFD-MKSLLTMGTPFNFEES-----NTSN--------HTQMLKDLISNKGNIPSSLMVY 226
             F  +   +++   +N E +     + SN         T+M + +  N   + S   +
Sbjct: 181 PNFPPVNKYVSLDGEYNNETNLQLGESLSNVLKEGPIVKTEMYQYIADNYQKVSSKTQML 240

Query: 227 NLAGT--NSYDGDKIVPFASVETGKYIFQETAKHYTQLTVTGNNATHSDLPDNPEVIQYV 284
            L G   +    D  +P+A   +  ++F++     T  T+     +HS  P NP V++YV
Sbjct: 241 LLEGNFNSEKQTDSAIPWADSFSIYHLFKKNGNEITT-TLYPTKTSHSQAPKNPTVVKYV 299
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8646 (GBS219) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 3; MW 31.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 7; MW 56kDa).

GBS219-GST was purified as shown in Figure 203, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 720

A DNA sequence (GBSx0764) was identified in *S.agalactiae* <SEQ ID 2213> which encodes the amino acid sequence <SEQ ID 2214>. This protein is predicted to be PTS system, cellobiose-specific IIC component. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.64    Transmembrane  263 - 279 ( 260 - 282)
   INTEGRAL    Likelihood = -6.26    Transmembrane  200 - 216 ( 197 - 226)
   INTEGRAL    Likelihood = -5.95    Transmembrane  157 - 173 ( 156 - 175)
   INTEGRAL    Likelihood = -5.79    Transmembrane  307 - 323 ( 306 - 332)
   INTEGRAL    Likelihood = -5.68    Transmembrane  131 - 147 ( 126 - 148)
   INTEGRAL    Likelihood = -4.73    Transmembrane  375 - 391 ( 370 - 396)
   INTEGRAL    Likelihood = -3.61    Transmembrane  101 - 117 (  98 - 119)
   INTEGRAL    Likelihood = -1.75    Transmembrane  326 - 342 ( 324 - 342)
   INTEGRAL    Likelihood = -0.37    Transmembrane   25 -  41 (  25 -  41)
   INTEGRAL    Likelihood = -0.16    Transmembrane   71 -  87 (  71 -  88)

----- Final Results -----
            bacterial membrane --- Certainty=0.4057(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC74807 GB:AE000268 PEP-dependent phosphotransferase enzyme II
           for cellobiose, arbutin, and salicin [Escherichia coli K12]
 Identities = 60/197 (30%), Positives = 83/197 (41%), Gaps = 12/197 (6%)

Query: 209 LAIFLTLSGLFVPDIL--FRPYSYFSVVSENLNAALSQHTDKIPYLYTFYTVKNSFAMFG 266
           LA+    +G+  P  L     Y  +  V   L A  + H    P L       +SF   G
Sbjct: 253 LALTALDNGIMTPWALENIATYQQYGSVEAALAAGKTFHIWAKPML-------DSFIFLG 305

Query: 267 GIGILLSLFLAVLYESRKLQSKNYYKLTLLTLTPLIFDQNLPFLVGLPVILQPILFIPMV 326
           G G  L L LA+   SR+    +Y ++  L L   IF  N P L GLP+I+ P++FIP V
Sbjct: 306 GSGATLGLILAIFIASRRA---DYRQVAKLALPSGIFQINEPILFGLPIIMNPVMFIPFV 362

Query: 327 LTTIFAEAFGALMLYLKFVDPAVYTVPSGTPSLLFGFLASNGDWRYLPVTAIILVVGFFI 386
           L      A      Y+  + P     P   P+ L  F  +NG    L V    L +   I
Sbjct: 363 LVQPILAAITLAAYYMGIIPPVTNIAPWTMPTGLGAFFNTNGSVAALLVALFNLGIATLI 422

Query: 387 YRPFVKIAFAKEEQYEK 403
           Y PFV +A   +   +K
Sbjct: 423 YLPFVVVANKAQNAIDK 439
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 721

A DNA sequence (GBSx0765) was identified in *S.agalactiae* <SEQ ID 2217> which encodes the amino acid sequence <SEQ ID 2218>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1991(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 722

A DNA sequence (GBSx0766) was identified in *S.agalactiae* <SEQ ID 2219> which encodes the amino acid sequence <SEQ ID 2220>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -5.79   Transmembrane  188 - 204 ( 179 - 206)
   INTEGRAL    Likelihood = -5.36   Transmembrane  105 - 121 ( 104 - 127)
   INTEGRAL    Likelihood = -4.41   Transmembrane  212 - 228 ( 210 - 229)
   INTEGRAL    Likelihood = -3.45   Transmembrane   72 -  88 (  69 -  89)
   INTEGRAL    Likelihood = -0.48   Transmembrane  124 - 140 ( 124 - 140)

----- Final Results -----
               bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8647> which encodes amino acid sequence <SEQ ID 8648> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 6
SRCFLG: 0
McG: Length  of  UR:   5
     Peak Value of UR:   2.99
     Net Charge of CR: 4
McG: Discrim Score:      6.88
GvH: Signal Score (-7.5): -2.86
     Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 5 value:  -5.79 threshold:  0.0
   INTEGRAL    Likelihood = -5.79   Transmembrane  179 - 195 ( 170 - 197)
   INTEGRAL    Likelihood = -5.36   Transmembrane   96 - 112 (  95 - 118)
   INTEGRAL    Likelihood = -4.41   Transmembrane  203 - 219 ( 201 - 220)
   INTEGRAL    Likelihood = -3.45   Transmembrane   63 -  79 (  60 -  80)
   PERIPHERAL  Likelihood =  0.10      18
 modified ALOM score:   1.66
icml HYPID: 7  CFP: 0.331

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
```

```
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2221> which encodes the amino acid sequence <SEQ ID 2222>. Analysis of this protein sequence reveals the following:

```
     Possible site: 30

>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-11.20    Transmembrane  179 - 195 ( 173 - 201)
   INTEGRAL    Likelihood = -3.66    Transmembrane   96 - 112 (  95 - 113)
   INTEGRAL    Likelihood = -1.44    Transmembrane  203 - 219 ( 203 - 219)
   INTEGRAL    Likelihood = -0.96    Transmembrane  115 - 131 ( 115 - 131)
   INTEGRAL    Likelihood = -0.64    Transmembrane   63 -  79 (  63 -  79)

----- Final Results -----
               bacterial membrane --- Certainty=0.5479(Affirmative) < succ>

                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 160/228 (70%), Positives = 185/228 (80%)

Query: 10  MSKKSHRQYQIYEGLRCAVALCFISGYINAFTYVTQGKRFAGVQTGNLLSFAIHLSNKHY 69
            MSKK  + YQ+YEGLRCA+ LCFISGY+NAFTY+TQGKRFAGVQTGNLLSFAI LS +
Sbjct: 1   MSKKKRKHYQVYEGLRCAMTLCFISGYVNAFTYMTQGKRFAGVQTGNLLSFAIRLSEQQL 60

Query: 70  SQALAFLLPIMVFMLGQSFTYFMNRWANKHQLHWYLLSSFALTQVAIVTIILTPFLPSSF 129
             +AL FLLP++VFMLGQSFTYFM+RWA K  LHWYLLSS  LT +A  T + TPFLPS+
Sbjct: 61  KEALQFLLPMIVFMLGQSFTYFMHRWATKKGLHWYLLSSVILTGIAFGTALFTPFLPSNV 120

Query: 130 TVAGLAFFASIQVDTFKSLRGAPYANMMMTGNIKNAAYLLTKGLYEKNSDIFLIARNTII 189
            TVA LAFFASIQVDTFK+LRGA YAN+MMTGNIKNAAYLLTKGLYEKN ++  I RNT+I
Sbjct: 121 TVAALAFFASIQVDTFKTLRGASYANVMMTGNIKNAAYLLTKGLYEKNHELTHIGRNTLI 180

Query: 190 IIGGFIFGVVCSTYFSSKLGEWSLSLILIPLLYVNLLLGHEFYNLQVE 237
            +I  F  GVVCST      GE++L  IL+PLLYVN LL  EFY++Q +
Sbjct: 181 VILAFAVGVVCSTLLCIAYGEYALMPILMPLLYVNYLLAQEFYHIQTK 228
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 723

A DNA sequence (GBSx0767) was identified in *S.agalactiae* <SEQ ID 2223> which encodes the amino acid sequence <SEQ ID 2224>. This protein is predicted to be tellurite resistance protein. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.00    Transmembrane  190 - 206 ( 190 - 206)

----- Final Results -----
               bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC22923 GB:U32807 tellurite resistance protein (tehB)
          [Haemophilus influenzae Rd]
 Identities = 164/282 (58%), Positives = 205/282 (72%), Gaps = 1/282 (0%)

Query: 7   LLPYKTMPVWTAQSIPKAFLEKHNTKEGTWAKLTILSGSLVFYQLSPDGEEISRHIFDAS 66
           L+ YK MPVWT  ++P+ F EKHNTK GTW KLT+L G L FY+L+ +G+ I+ HIF
Sbjct: 5   LICYKQMPVWTKDNLPQMFQEKHNTKVGTWGKLTVLKGKLKFYELTENGDVIAEHIFTPE 64

Query: 67  SDIPFVDPQVWHKVSPNSPDLSCYLTFYCQKEDYFHKKYGLTRTHSEVIASAPLLSEKSN 126
           S IPFV+PQ WH+V   S DL C L FYC+KEDYF KKY  T  H +V+ +A ++S
Sbjct: 65  SHIPFVEPQAWHRVEALSDDLECTLGFYCKKEDYFSKKYNTTAIHGDVVDAAKIISP-CK 123

Query: 127 ILDLGCGQGRNSLYLSLLGHQVTSVDSNGQSLVALENMALEEELPYNIKRYDINTAAIEG 186
           +LDLGCGQGRNSLYLSLLG+ VTS D N  S+  L     +E L +    YDIN A I+
Sbjct: 124 VLDLGCGQGRNSLYLSLLGYDVTSWDHNENSIAFLNETKEKENLNISTALYDINAANIQE 183

Query: 187 HYDFILSTVVFMFLNPDCISDIILQMQSHTQIGGYNLIVSAMDTAENPCPLPFPFTFKEG 246
           +YDFI+STVVFMFLN + +  II  M+ HT +GGYNLIV+AM T + PCPLPF FTF E
Sbjct: 184 NYDFIVSTVVFMFLNRERVPSIIKNMKEHTNVGGYNLIVAAMSTDDVPCPLPFSFTFAEN 243

Query: 247 QLKSYYNDWEIIKYNENLGELHRVDENGNRLKLQFATLLARK 288
           +LK YY DWE ++YNEN+GELH+ DENGNR+K++FAT+LARK
Sbjct: 244 ELKEYYKDWEFLEYNENMGELHKTDENGNRIKMKFATMLARK 285
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 2224 (GBS95) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 3; MW 35.6kDa) and in Figure 12 (lane 4; MW 35.6kDa). The GBS95-His fusion product was purified (Figure 191, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 292), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 724

A DNA sequence (GBSx0768) was identified in *S.agalactiae* <SEQ ID 2225> which encodes the amino acid sequence <SEQ ID 2226>. This protein is predicted to be methionyl-tRNA synthetase (metS). Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane  473 - 489 ( 473 - 489)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10043> which encodes amino acid sequence <SEQ ID 10044> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB11814 GB:Z99104 methionyl-tRNA synthetase [Bacillus subtilis]
 Identities = 395/667 (59%), Positives = 501/667 (74%), Gaps = 12/667 (1%)

Query: 20  EKKSFYITTPIYYPSGKLHIGSAYTTIACDVLARYKRMMGFDVQYLTGLDEHGQKIQQKA 79
           E  +FYITTPIYYPSGKLHIG AYTT+A D +ARYKR+ GFDV+YLTG DEHGQKIQQKA
Sbjct: 4   ENNTFYITTPIYYPSGKLHIGHAYTTVAGDAMARYKRLKGFDVRYLTGTDEHGQKIQQKA 63
```

```
Query: 80  EEAGITPQEYVDGMAESVKTLWELLDISYDKFIRTTDTYHEEAVAKIFEQLLAQGDIYLG 139
           E+  ITPQEYVD  A  ++ LW+ L+IS D FIRTT+  H+  + K+F++LL  GDIYL
Sbjct: 64  EQENITPQEYVDRAAADIQKLWKQLEISNDDFIRTTEKRHKVVIEKVFQKLLDNGDIYLD 123

Query: 140 EYTGWYSVSDEEFFTESQLAEVYRDENGNMIGGVAP-SGHEVEKVSEESYFFRMSKYADR 198
           EY GWYS+ DE F+TE+QL ++ R+E G +IGG +P SGH VE + EESYFFRM KYADR
Sbjct: 124 EYEGWYSIPDETFYTETQLVDIERNEKGEVIGGKSPDSGHPVELIKEESYFFRMGKYADR 183

Query: 199 LKAYYAEHPEFIQPDGRMNEMLKNFIEPGLEDLAVSRTTYTWGVQVPSNPKHVIYVWIDA 258
           L  YY E+P FIQP+ R NEM+ NFI+PGLEDLAVSRTT+ WGV+VP NPKHV+YVWIDA
Sbjct: 184 LLKYYEENPTFIQPESRKNEMINNFIKPGLEDLAVSRTTFDWGVKVPENPKHVVYVWIDA 243

Query: 259 LMNYISALGYGWSDDLSQYHKFWPADIHMIGKDILRFHSIYWPIMLMALDLPLPKRLVAH 318
           L NY++ALGY   +D   Y K+WPAD+H++GK+I+RFH+IYWPIMLMALDLPLPK++ AH
Sbjct: 244 LFNYLTALGYDTEND-ELYQKYWPADVHLVGKEIVRFHTIYWPIMLMALDLPLPKQVFAH 302

Query: 319 GWFVMQDGKMSKSKGNVVYPEMLVERFGLDPLRYYLMRSLPVGSDGTFTPEDYVGRINYE 378
           GW +M+DGKMSKSKGNVV P  L+ER+GLD LRYYL+R +P GSDG FTPE +V RINY+
Sbjct: 303 GWLLMKDGKMSKSKGNVVDPVTLIERYGLDELRYYLLREVPFGSDGVFTPEGFVERINYD 362

Query: 379 LANDLGNLLNRTIAMVNKYFDGEVPRF-AVATDFDADLASVATDSIENYHKQMEAVDFPR 437
           LANDLGNLLNRT+AM+NKYFDG++  +    T+FD  L SVA ++++ Y K ME ++F
Sbjct: 363 LANDLGNLLNRTVAMINKYFDGQIGSYKGAVTEFDHTLTSVAEETVKAYEKAMENMEFSV 422

Query: 438 ALEAVWNLISRTNKYIDETAPWVLAKDETDRDKLAAVMSHLVASLRVVAHLIQPFMMETS 497
           AL  +W LISRTNKYIDETAPWVLAKD    ++L +VM HL  SLR+ A L+QPF+ +T
Sbjct: 423 ALSTLWQLISRTNKYIDETAPWVLAKDPAKEEELRSVMYHLAESLRISAVLLQPFLTKTP 482

Query: 498 DAIMEQLGL--GATFDLEKLT-FADLPEGVRVVAKGSPIFPRLDMEDEITYIKEQMNAGK 554
           + + EQLG+   +    + +T F  L +    V KG P+FPRL+ E+EI YIK +M  G
Sbjct: 483 EKMFEQLGITDESLKAWDSITAFGQLKD--TKVQKGEPLFPRLEAEEEIAYIKGKMQ-GS 539

Query: 555 APVEKEWVPEEVELTSSKGQIKFEDFDAVEIRVAEVIEVEKVEGSDKLLRFRLDAGDEGH 614
           AP ++E   EE +      +I  + F  VE+RVAEVIE E V+ +D+LL+ +LD G E
Sbjct: 540 APAKEETKEEEPQEVDRLPEITIDQFMDVELRVAEVIEAEPVKKADRLLKLQLDLGFE-K 598

Query: 615 RQILSGIAKFYPNEQELVGKKLQIVANLKPRKMMKKYVSQGMILSAEHDGKLTVLTVDSA 674
           RQ++SGIAK Y  E ELVGKKL  V NLKP K ++  +SQGMIL+ E DG L V+++D +
Sbjct: 599 RQVVSGIAKHYTPE-ELVGKKLVCVTNLKPVK-LRGELSQGMILAGEADGVLKVVSIDQS 656

Query: 675 VANGSII 681
           +  G+ I
Sbjct: 657 LPKGTRI 663
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2227> which encodes the amino acid sequence <SEQ ID 2228>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1245(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 516/665 (77%), Positives = 573/665 (85%), Gaps = 4/665 (0%)

Query: 21  KKSFYITTPIYYPSGKLHIGSAYTTIACDVLARYKRMMGFDVQYLTGLDEHGQKIQQKAE 80
           KK FYITTPIYYPSGKLHIGSAYTTIACDVLARYKR+MG +V YLTGLDEHGQKIQ KA+
Sbjct: 3   KKPFYITTPIYYPSGKLHIGSAYTTIACDVLARYKRLMGHEVFYLTGLDEHGQKIQTKAK 62

Query: 81  EAGITPQEYVDGMAESVKTLWELLDISYDKFIRTTDTYHEEAVAKIFEQLLAQGDIYLGE 140
           EAGITPQ YVD MA+ VK LW+LLDISYD FIRTTD YHEE VA +FE+LLAQ DIYLGE
Sbjct: 63  EAGITPQTYVDNMAKDVKALWQLLDISYDTFIRTTDDYHEEVVAAVFEKLLAQDDIYLGE 122
```

```
Query: 141 YTGWYSVSDEEFFTESQLAEVYRDENGNMIGGVAPSGHEVEKVSEESYFFRMSKYADRLK 200
           Y+GWYSVSDEEFFTESQL EV+RDE+G +IGG+APSGHEVE VSEESYF R+SKY DRL
Sbjct: 123 YSGWYSVSDEEFFTESQLKEVFRDEDGQVIGGIAPSGHEVEWVSEESYFLRLSKYDDRLV 182

Query: 201 AYYAEHPEFIQPDGRMNEMLKNFIEPGLEDLAVSRTTYTWGVQVPSNPKHVIYVWIDALM 260
           A++ E P+FIQPDGRMNEM+KNFIEPGLEDLAVSRTT+TWGV VPS+PKHV+YVWIDAL+
Sbjct: 183 AFFKERPDFIQPDGRMNEMVKNFIEPGLEDLAVSRTTFTWGVPVPSDPKHVVYVWIDALL 242

Query: 261 NYISALGYGWSDDLSQYHKFWPADI-HMIGKDILRFHSIYWPIMLMALDLPLPKRLVAHG 319
           NY +ALGY  ++  + + KFW   + HM+GKDILRFHSIYWPI+LM LDLP+P RL+AHG
Sbjct: 243 NYATALGYRQANH-ANFDKFWNGTVFHMVGKDILRFHSIYWPILLMMLDLPMPDRLIAHG 301

Query: 320 WFVMQDGKMSKSKGNVVYPEMLVERFGLDPLRYYLMRSLPVGSDGTFTPEDYVGRINYEL 379
           WFVM+DGKMSKSKGNVVYPEMLVERFGLDPLRYYLMRSLPVGSDGTFTPEDYVGRINYEL
Sbjct: 302 WFVMKDGKMSKSKGNVVYPEMLVERFGLDPLRYYLMRSLPVGSDGTFTPEDYVGRINYEL 361

Query: 380 ANDLGNLLNRTIAMVNKYFDGEVPRFA-VATDFDADLASVATDSIENYHKQMEAVDFPRA 438
           ANDLGNLLNRT+AM+NKYFDG VP +    T FDADL+ +    + +YHK MEAVD+PRA
Sbjct: 362 ANDLGNLLNRTVAMINKYFDGTVPAYVDNGTAFDADLSQLIDAQLADYHKHMEAVDYPRA 421

Query: 439 LEAVWNLISRTNKYIDETAPWVLAKDETDRDKLAAVMSHLVASLRVVAHLIQPFMMETSD 498
           LEAVW +I+RTNKYIDETAPWVLAK++ D+ +LA+VM+HL ASLR+VAH+IQPFMMETS
Sbjct: 422 LEAVWTIIARTNKYIDETAPWVLAKEDGDKAQLASVMAHLAASLRLVAHVIQPFMMETSA 481

Query: 499 AIMEQLGLGATFDLEKLTFADLPEGVRVVAKGSPIFPRLDMEDEITYIKEQMNAGKA-PV 557
           AIM QLGL    DL  L  AD P   +VVAKG+PIFPRLDME EI YIK QM    A
Sbjct: 482 AIMAQLGLEPVSDLSTLALADFPANTKVVAKGTPIFPRLDMEAEIDYIKAQMGDSSAISQ 541

Query: 558 EKEWVPEEVELTSSKGQIKFEDFDAVEIRVAEVIEVEKVEGSDKLLRFRLDAGDEGHRQI 617
           EKEWVPEEV L S K  I FE FDAVEIRVAEV EV KVEGS+KLLRFR+DAGD   RQI
Sbjct: 542 EKEWVPEEVALKSEKDVITFETFDAVEIRVAEVKEVSKVEGSEKLLRFRVDAGDGQDRQI 601

Query: 618 LSGIAKFYPNEQELVGKKLQIVANLKPRKMMKKYVSQGMILSAEHDGKLTVLTVDSAVAN 677
           LSGIAKFYPNEQELVGKKLQIVANLKPRKMMKKY+SQGMILSAEH  +LTVLTVDS+V N
Sbjct: 602 LSGIAKFYPNEQELVGKKLQIVANLKPRKMMKKYISQGMILSAEHGDQLTVLTVDSSVPN 661

Query: 678 GSIIG 682
           GSIIG
Sbjct: 662 GSIIG 666
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 725

A DNA sequence (GBSx0769) was identified in *S.agalactiae* <SEQ ID 2229> which encodes the amino acid sequence <SEQ ID 2230>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2633(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes.*

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 726

A DNA sequence (GBSx0770) was identified in *S.agalactiae* <SEQ ID 2231> which encodes the amino acid sequence <SEQ ID 2232>. This protein is predicted to be branched chain amino acid transport system II carrier protein (brnQ). Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-14.91   Transmembrane  279 - 295 ( 269 - 303)
   INTEGRAL    Likelihood = -9.98   Transmembrane   82 -  98 (  74 - 102)
   INTEGRAL    Likelihood = -6.58   Transmembrane  345 - 361 ( 340 - 364)
   INTEGRAL    Likelihood = -6.00   Transmembrane  157 - 173 ( 153 - 179)
   INTEGRAL    Likelihood = -4.30   Transmembrane   48 -  64 (  45 -  66)
   INTEGRAL    Likelihood = -4.14   Transmembrane  251 - 267 ( 250 - 278)
   INTEGRAL    Likelihood = -4.09   Transmembrane  308 - 324 ( 305 - 326)
   INTEGRAL    Likelihood = -2.55   Transmembrane  218 - 234 ( 216 - 237)
   INTEGRAL    Likelihood = -1.38   Transmembrane  126 - 142 ( 126 - 142)

----- Final Results -----
               bacterial membrane --- Certainty=0.6965(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9407> which encodes amino acid sequence <SEQ ID 9408> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC00400 GB:AF008220 branch-chain amino acid transporter
           [Bacillus subtilis]
 Identities = 130/367 (35%), Positives = 204/367 (55%), Gaps = 12/367 (3%)

Query: 1   MSEKFSPWFSLTFLVILYLTIGPLFAIPRTATVSFEIGVAPIVGHSP--IALLCFTACFF 58
           +++K  P F   F V+LYL+IGPLFAIPRT TVS+EIG  P +   P  ++LL FT  FF
Sbjct: 73  LADKAHPVFGTIFTVVLYLSIGPLFAIPRTGTVSYEIGAVPFLTGVPERLSLLIFTLIFF 132

Query: 59  AAAYYLAIRPNGILDSVGKILTPVFAFLILSLVVVGAIAYGNLESAKASADYAGKAFGSG 118
              YYLA+ P+ ++D VGKILTP+  F I+ ++V+ AI          +  Y G     G
Sbjct: 133 GVTYYLALNPSKVVDRVGKILTPI-KFTIILIIVLKAIFTPMGGLGAVTEAYKGTPVFKG 191

Query: 119 VLAGYNTLDALAAVAFCLVATETLKKFGFKTKKEYLSTIWIVGIVTSLAFSILYIGLGFL 178
            L GY T+DALA++ F +V    +K  G    K   +     G++ +L  + +Y+ L +L
Sbjct: 192 FLEGYKTMDALASIVFGVVVVNAVKSKGVTQSKALAAACIKAGVIAALGLTFIYVSLAYL 251

Query: 179 GNKFPVPADILADPNVNKGAYVLSQASYKLFGNFGRYFLSIMVTLTCFTTTVGLIVSVSE 238
           G          A   V +GA +LS +S+ LFG+ G   L   +T+ C TT++GL+ S  +
Sbjct: 252 G-----ATSTNAIGPVGEGAKILSASSHYLFGSLGNIVLGAAITVACLTTSIGLVTSCGQ 306

Query: 239 FFDKNFRFGNYKLFATVFTLIGFLIANLGLNAVITFSVPVLTLLYPIVIVIVLIILINKW 298
           +F K     +YK+  T+ TL   +IAN GL  +I FSVP+L+ +YP+ IVI+++  I+K
Sbjct: 307 YFSKLIPALSYKIVVTIVTLFSLIIANFGLAQIIAFSVPILSAIYPLAIVIIVLSFIDKI 366

Query: 299 LPLSKK---GMSLTIGLVTLVSFVEVLAGQWQEKTLTQLVGFLPFHTISMGWLVPMLIGI 355
               ++      +  GL +++  ++  AG         L   LP +++ +GW++P ++G
Sbjct: 367 FKERREVYIACLIGTGLFSILDGIKA-AGFSLGSLDVFLNANLPLYSLGIGWVLPGIVGA 425

Query: 356 VFSLVLS 362
           V   VL+
Sbjct: 426 VIGYVLT 432
```

There is also homology to SEQ ID 2234.

A related GBS gene <SEQ ID 8649> and protein <SEQ ID 8650> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length  of  UR:   30
     Peak Value of UR:   2.99
     Net Charge of CR: 2
McG: Discrim Score:     13.17
GvH: Signal Score (-7.5): -3.3
     Possible site: 33
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 11 value: -14.91 threshold:  0.0
   INTEGRAL    Likelihood =-14.91   Transmembrane  347 - 363 ( 337 - 371)
   INTEGRAL    Likelihood = -9.98   Transmembrane  150 - 166 ( 142 - 170)
   INTEGRAL    Likelihood = -7.54   Transmembrane   40 -  56 (  36 -  61)
   INTEGRAL    Likelihood = -6.64   Transmembrane   79 -  95 (  76 -  97)
   INTEGRAL    Likelihood = -6.00   Transmembrane  225 - 241 ( 221 - 247)
   INTEGRAL    Likelihood = -4.30   Transmembrane  116 - 132 ( 113 - 134)
   INTEGRAL    Likelihood = -4.14   Transmembrane  319 - 335 ( 318 - 346)
   INTEGRAL    Likelihood = -4.09   Transmembrane  376 - 392 ( 373 - 394)
   INTEGRAL    Likelihood = -2.92   Transmembrane    7 -  23 (   6 -  28)
   INTEGRAL    Likelihood = -2.55   Transmembrane  286 - 302 ( 284 - 305)
   INTEGRAL    Likelihood = -1.38   Transmembrane  194 - 210 ( 194 - 210)
   PERIPHERAL  Likelihood =  2.49     402
 modified ALOM  score:   3.48
icml HYPID: 7  CFP: 0.696

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6965(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00247(304 - 1596 of 1941)
OMNI|NT01BS3447(19 - 446 of 459) branched chain amino acid transport system II carrier
protein
%Match = 21.7
%Identity = 38.8  %Similarity = 61.2
Matches = 166  Mismatches = 157  Conservative Sub.s = 96

93        123       153       183       213       243       273       303
VLTVDSAVANGSIIG*SKRALCSFFVFKKKVTE*LENYENDLEFIFIFDIIKDIDSKHLDRI**GEFMERV*IDYLH*WL

                                                              LTEYFNIIIRRIFFMKHS
                                                                      10

333       363       393       423       453       483       513       543
LMVKKGFLTGLLLFGIFFGAGNLIFPPALGVASGQDFWPAILGFCLSGVGLAIITLLLGTLTNGGYKTEMSEKFSPWFSL
| ||   : |::|| :|||||||:|:|| || |:|:: | || || |:||||| :: ::   ||    |  :::|  | |  
LPVKDTIIIGFMLFALFFGAGNMIYPPELGQAAGHNVWKAIGGFLLTGVGLPLLGIIAIALTGKDAKG-LADKAHPVFGT
          30        40        50        60        70        80        90

573       603       633       657       687       717       747       777
TFLVILYLTIGPLFAIPRTATVSFEIGVAPIVGHSPIALLCFTACFFAAAYYLAIRPNGILDSVGKILTPVFAFLILS
 | |:|||:||||||||||| |||:|||  | :   |  ::|| ||  ||     ||||: |: ::| |||||||:  | |: 
IFTVVLYLSIGPLFAIPRTGTVSYEIGAVPFLTGVPERLSLLIFTLIFFGVTYYLALNPSKVVDRVGKILTPI-KFTIIL
          110       120       130       140       150       160       170

801       831       861       891       921       951       981       1011
LVVVGAI--AYGNLESAKASADYAGKAFGSGVLAGYNTLDALAAVAFCLVATETLKKFGFKTKKEYLSTIWIVGIVTSLA
::|: ||    | | :   :  | |     | | || |:||||:: | :|    :|  |    |   :     |:: :| 
IIVLKAIFTPMGGLGA--VTEAYKGTPVFKGFLEGYKTMDALASIVFGVVVVNAVKSKGVTQSKALAAACIKAGVIAALG
           190         200       210       220       230       240       250

1041      1071      1101      1131      1161      1191      1221      1251
FSILYIGLGFLGNKFPVPADILADPNVNKGAYVLSQASYKLFGNFGRYFLSIMVTLTCFTTTVGLIVSVSEFFDKNFRFG
```

```
:: :|: | :||          |   | :|| :|| :|: |||::|   |   :|: |:||::||: |  ::| |
LTFIYVSLAYLG-----ATSTNAIGPVGEGAKILSASSHYLFGSLGNIVLGAAITVACLTTSIGLVTSCGQYFSKLIPAL
                  270       280       290       300       310       320

1281      1311      1341      1371      1401      1431      1461      1488
NYKLFATVFTLIGFLIANLGLNAVITFSVPVLTLLYPIVIVIVLIILINKWLPLSKKGMSLTIGLVTLVSFVEVLAG-QW
:||:  |: ||  ::|||:||  :| ||||:|: :||: |||::: :|:|   : |:   : |  :    :  :| | :
SYKIVVTIVTLFSLIIANFGLAQIIAFSVPILSAIYPLAIVIIVLSFIDK---IFKERREVYIACLIGTGLFSILDGIKA
        340       350       360       370          380       390       400

1518      1536      1566      1596      1626      1656      1686      1716
QEKTLTQLVGFL----PFHTISMGWLVPMLIGIVFSLVLSDKQKGQAFDLEKFEG*HYFNFIDMSKRLKLRF*PFLYQIF
   :|  |  ||    |:::: :||::| ::| |   ||:
AGFSLGSLDVFLNANLPLYSLGIGWVLPGIVGAVIGYVLTLFIGPSKQLNEIS
           420       430       440       450
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 727

A DNA sequence (GBSx0771) was identified in *S.agalactiae* <SEQ ID 2235> which encodes the amino acid sequence <SEQ ID 2236>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3291(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10041> which encodes amino acid sequence <SEQ ID 10042> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 728

A DNA sequence (GBSx0772) was identified in *S.agalactiae* <SEQ ID 2237> which encodes the amino acid sequence <SEQ ID 2238>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.33   Transmembrane  117 - 133 ( 112 - 136)
   INTEGRAL    Likelihood = -3.77   Transmembrane   53 -  69 (  53 -  70)
   INTEGRAL    Likelihood = -3.40   Transmembrane   98 - 114 (  97 - 115)

----- Final Results -----
               bacterial membrane --- Certainty=0.4333(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 729

A DNA sequence (GBSx0773) was identified in *S.agalactiae* <SEQ ID 2239> which encodes the amino acid sequence <SEQ ID 2240>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.19   Transmembrane   22 -  38 (  20 -  44)

----- Final Results -----
               bacterial membrane --- Certainty=0.2678(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8651> which encodes amino acid sequence <SEQ ID 8652> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length  of  UR:   21
     Peak Value of UR:   3.11
     Net Charge of CR: 2
McG: Discrim Score:     11.30
GvH: Signal Score (-7.5): -5.35
     Possible site: 28
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 1 value:  -4.19 threshold:  0.0
   INTEGRAL    Likelihood = -4.19   Transmembrane    5 -  21 (   3 -  27)
   PERIPHERAL  Likelihood =  6.74      53
 modified ALOM score:   1.34
icm1 HYPID: 7  CFP: 0.268

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2678(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15623 GB:Z99122 spore coat protein (inner) [Bacillus subtilis]
 Identities = 71/359 (19%), Positives = 148/359 (40%), Gaps = 49/359 (13%)

Query: 127 ISYRGNTSRYFDKKSLKVKFVTNKLKEKKHRLAGMPKESEWVLHGPFLDRTLLRNYLSYN 186
           I+YRG+  R F KKS  + F   K       +          L+  + D +L+RN LS +
Sbjct: 47  IAYRGSHIRDFKKKSYHISFYQPKTFRGAREIH---------LNAEYKDPSLMRNKLSLD 97

Query: 187 IAGEIMSYAPNVRYCELFVNGEYQGVYLAVENIEQGEQRVPIEKSDKKLHKTPYIVAWDR 246
              E+ + +P   +  + +NG+ +GVYL +E++++         + +KL       A D
Sbjct: 98  FFSELGTLSPKAEFAFVKMNGKNEGVYLELESVDE------YYLAKRKLADGAIFYAVDD 151

Query: 247 EHKAKQKLDNYVHYTHQSGISALDVKYPGKQRLTSKQLEFINKD----INHIEKVLYSYD 302
           +       D       +   ++L++ Y  +++  +++ +F  +D    IN + K  +
Sbjct: 152 DANFSLMSD-----LERETKTSLELGY--EKKTGTEEDDFYLQDMIFKINTVPKAQFK-- 202

Query: 303 FSQYPKYIDRESFANYFVINEFFRNVDAGKFSTYLYKDLRDRA-KLVVWDFNNAFDNQIE 361
            S+  K++D + +  +     F  N D    +  LY+       +++ WD++  +   I
Sbjct: 203 -SEVTKHVDVDKYLRWLAGIVFTSNYDGFVHNYALYRSGETGLFEVIPWDYDATWGRDIH 261

Query: 362 GRVDEADFTLTDAPWFNMLIKDKAFIDLVVHRYKELRKGVLATEYLSNYIDETRHFLGPA 421
```

```
           G    AD+       FN L             YK L +  L + +   Y++       P
Sbjct: 262 GERMAADYVRIQG--FNTLTARILDESEFRKSYKRLLEKTLQSLFTIEYME-------PK 312

Query: 422 IDRNYKKWGYVFDLKNTDPRNYLIPTERN-VTSYHKSVEQLKDFIKKRGRWMDRNIETL 479
           I   Y++           P   + P ++N +  + +  + + ++IK R +++  ++  L
Sbjct: 313 IMAMYER---------IRPFVLMDPYKKNDIERFDREPDVICEYIKNRSQYLKDHLSIL 362
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 730

A DNA sequence (GBSx0774) was identified in *S.agalactiae* <SEQ ID 2241> which encodes the amino acid sequence <SEQ ID 2242>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 731

A DNA sequence (GBSx0775) was identified in *S.agalactiae* <SEQ ID 2243> which encodes the amino acid sequence <SEQ ID 2244>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -4.62    Transmembrane    5 -  21 (   3 -  24)

----- Final Results -----
               bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05949 GB:AP001514 unknown [Bacillus halodurans]
 Identities = 199/697 (28%), Positives = 322/697 (45%), Gaps = 58/697 (8%)

Query: 57   KPFVVKGVDVESSLAGYHHNDFPITQKTYREWFHLISNMGANTVRVKVPMNVAFYDALYH 116
            K   + GV++     G    +  I +K Y  WF  I  MG N +RV       FY AL
Sbjct: 414  KKLQIHGVNLGMGKPGTFPGEAAIKEKDYYRWFEQIGEMGGNAIRVYTLHPPGFYHALKR 473

Query: 117  HNKASKRPLYLLQGIRIDSYRNNASITAFNDNYRGYLKREAKGVVDILHGRKQVWNTDLG 176
            +N+  + P+YL  G+ ID      ++ AF++      ++E K +VD++HG   V + + G
Sbjct: 474  YNEQHENPIYLFHGVWIDEEPLEDTLDAFDEETNEEFQQEMKRIVDVIHGNAVV-DPNPG 532

Query: 177  SRH--YHYDLSPWVLGYVVGDDWNSGTVAYTNHQEKKT-QYKGRYFKTSVAANPFEVMLA 233
              H  Y  D+SP+ +G+++G +W   TV  TN        Y G+Y +T   A PFE  LA
Sbjct: 533  HAHGVYQADVSPYTIGWIIGIEWYPHTVKATNKNNPDIGDYDGKYVETK-DAEPFEYWLA 591
```

```
Query: 234  QVMDELTHYETAKYGWQHLISFSNSPTTDPF-HYRKPFEAQAPKYVQLNVENIQANSNVK 292
               D L  YE  +Y W   +SF+N  TTD   H  +P E +    V  NV +++  +  +
Sbjct: 592  NQFDILLSYEIEQYNWIRPVSFTNWVTTDLLTHPAEPNEDEDLVGVDPNVIHLKGPA-TE 650

Query: 293  AGMFAAYKAIDFHPRYKDYLLFDKENISKEDRQKIKELSLSQGYVKLLNAYHKIPVLVTG 352
               FA+Y     +P Y D+L ++++ I   D +   EL+   GY+K L+  H +P+L+
Sbjct: 651  TNQFASYHV---YPYYPDFLNYEEDYIHYVDHR--GELNNYAGYLKDLHDAHDLPILIAE 705

Query: 353  YGYSTARGIA-QKEIDKRPLPINEKEQGQRLLEDYESFISSGSFGATINAWQDDWNARAW 411
            +G   +RG+  +    K    ++E+EQG+ ++E +E  I     G  I  WQD+W  R W
Sbjct: 706  FGVPASRGLTHENPFGKNQGFLSEEEQGKIVVELFEDIIEEKLLGGLIFTWQDEWFKRTW 765

Query: 412  NTSFATNKHSQFLWGDAQVFNQGYGLLGFKNAKHHYQVDGKRGKG-----EWKHPLMTSA 466
            NT    N   +  W +AQ   Q +GLL F   K     D +  +      E  HP +
Sbjct: 766  NTMDYDNPDRRPFWSNAQTNEQQFGLLSFDRLKVKVNGDDQDWEDASLLYEEDHPYVKR- 824

Query: 467  TGDDLYASSDESYLYLAIKTKPEKLKE-----KRLLPIDITPKSGSRKMNGSK-VTFSKS 520
                LY   DE YLY  I  K     +       +L +D  P  G+  +   + VTF
Sbjct: 825  ----LYMDHDERYLYFRIDMKSGSTDDFFKDGFPILVLDTLPGQGNEHIKEVEGVTFDHG 880

Query: 521  SDFVLSIDPNGKSELFVQERYNALKANYLRQLNGKDFYAFPPKKNSSNFEQINMVLRNTK 580
             DF++ +    +S + V   Y+     Y +     +  +  P+ N+  F++I+  L N +
Sbjct: 881  IDFIIELKGYDESRVKVDAYYDFFTYQYSQIYQMIEETSIEPQNNTGVFQKIHYAL-NQE 939

Query: 581  IVEDMEKVKATERFLP--THPTGLLKTGTTDRHQKTFDSQTD--ISFGKDFIEVRIPWQL 636
            I     ++ +T   +P   + TG L+ G  D     +DS  D  ++  K  IEVRIPW L
Sbjct: 940  I-----RIPSTNEVIPFSYYETGELRHGNGDPEADDYDSLADFFVNEEKGMIEVRIPWLL 994

Query: 637  LNFSDPSSQKIHDDYFKHYGVKELE-IESI-ALGLGANSKENTLIKMAD----------- 683
            L+F DPS +++    ++  G +  E IE + A  L    K++   ++ D
Sbjct: 995  LSFKDPSQREVMSAIYEGEGGETSEIIEGVRAAVLFVEPKDDDSYQVVDALPALDGDRLT 1054

Query: 684  ------YRLKNWERPDTKTFLKDSYYSIKKEWSKERE 714
                  Y  + W+ P  +  LK SY  +K+ ++  +E
Sbjct: 1055 DEVMNMYTWETWDIPLYEERLKQSYDLVKEAFTSIKE 1091
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8653> and protein <SEQ ID 8654> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 2
McG: Discrim Score:      12.00
GvH: Signal Score (-7.5): -5.46
     Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -4.62 threshold:  0.0
   INTEGRAL    Likelihood = -4.62   Transmembrane    5 -  21 (   3 -  24)
   PERIPHERAL  Likelihood =  7.32     223
 modified ALOM score:   1.42

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 2244 (GBS62) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 7; MW 80.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 4; MW 105kDa).

The GBS62-GST fusion product was purified (Figure 100A; see also Figure 193, lane 7) and used to immunise mice (lane 1 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure

100B), FACS (Figure 100C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 732

A DNA sequence (GBSx0778) was identified in *S.agalactiae* <SEQ ID 2245> which encodes the amino acid sequence <SEQ ID 2246> in others. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.48    Transmembrane  310 - 326 ( 302 - 335)
   INTEGRAL    Likelihood = -7.32    Transmembrane  362 - 378 ( 361 - 380)
   INTEGRAL    Likelihood = -7.11    Transmembrane  334 - 350 ( 329 - 355)
   INTEGRAL    Likelihood = -2.28    Transmembrane  381 - 397 ( 380 - 397)

----- Final Results -----
               bacterial membrane --- Certainty=0.3994(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10039> which encodes amino acid sequence <SEQ ID 10040> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05950 GB:AP001514 unknown conserved protein in others
          [Bacillus halodurans]
 Identities = 143/405 (35%), Positives = 226/405 (55%), Gaps = 5/405 (1%)

Query: 11  IVPAYNESTTIVSSIDSLLHLDYEAYEIIVVDDGSSDNTSDVLKEEFALMKISNTIDSII 70
           +VPAYNE T I+ ++ SLL L Y   EI+VV+DGS+D T +V+ E F ++K+   I   I
Sbjct: 69  LVPAYNEETGIIETVRSLLSLKYPQTEIVVVNDGSTDQTLEVIIEHFQMVKVGKVIRKQI 128

Query: 71  ATQTCKDVFQRQVGKVKLTLIVKENGGKGDALNMGINAANYDYFLCLDADSMLQVDSLSQ 130
            T+  K V+Q  +    L L+ K NGGK DALN G+N + Y YF  +D DS+L+ D+L +
Sbjct: 129 ETEPIKGVYQSTIFP-HLLLVDKSNGGKADALNAGLNVSKYPYFCSIDGDSILETDALLK 187

Query: 131 ISKSIQV----DPTVIAVGGLVQVAQGVKIEQGKVASYRLPWRIIPCAQALEYDSSFLGA 186
           + K I      +  VIA GG V++A G  I+ G V S +L    +   Q +EY  +FL
Sbjct: 188 VMKPIVTSRDDEDEVIASGGNVRIANGSDIQMGSVLSVQLAKNPLVVMQVIEYLRAFLMG 247

Query: 187 RIFLDYLRANLIISGAFGLFKKDLVKAVGGYDTQTLGEDMELVMKLHFFCRNNNIPYRIC 246
           RI L      LIISGAF +F K  V   GGY  +T+GEDMELV++LH   +   +  RI
Sbjct: 248 RIGLSRHNMVLIISGAFSVFAKKWVMEAGGYSKKTVGEDMELVVRLHRLVKEKRLKKRIT 307

Query: 247 YETDAVCWSQAPTNLGDLRKQRRRWYLGLYQCLKKYKSIFANYRFGAVGSISYIYYILFE 306
           +  D VCW++AP     L++QR RW+ GL + L  ++ +  N ++G VG+ S  Y+ + E
Sbjct: 308 FVPDPVCWTEAPATFRVLQRQRSRWHRGLMESLWLHRGMTFNPKYGLVGTASIPYFWIVE 367

Query: 307 LLTPFIECFGIVIIFLSLLFNQLNIPFFISLVSLYIFYCVLITLSSFLHRIYSQQLVIGI 366
              P +E  G + I  +  F  L + F ++L  L++ Y  + ++++ +   +S +    +
Sbjct: 368 FFGPVVELMGYLYIVFAFFFGGLYVEFALALFLLFVLYGTVFSMTAVILEGWSLKRYPKV 427

Query: 367 LDIVKVFYIAVFRYLILHPVLTFVKVASVIGYKNKKMVWGHITRE 411
            D+ ++   ++F  L   P+    +  ++I    +   WG +TR+
Sbjct: 428 SDMSRLMIFSLFEALWYRPLTVLWRFGAIIEALFRSKAWGEMTRK 472
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2247> which encodes the amino acid sequence <SEQ ID 2248>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.04   Transmembrane    33 -  49 (  24 -  57)
   INTEGRAL    Likelihood =-10.77   Transmembrane   376 - 392 ( 370 - 399)
   INTEGRAL    Likelihood = -7.86   Transmembrane   344 - 360 ( 342 - 372)
   INTEGRAL    Likelihood = -4.94   Transmembrane    63 -  79 (  55 -  81)
   INTEGRAL    Likelihood = -2.07   Transmembrane   403 - 419 ( 403 - 419)

----- Final Results -----
               bacterial membrane --- Certainty=0.5416(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 84/397 (21%), Positives = 173/397 (43%), Gaps = 71/397 (17%)

Query: 6    FRRKSIVPAYNEST-TIVSSIDSLLHLDYEAYEIIVVDDGSSDNTSDVLKEEFALMKISN 64
            ++  +++P+YNE   +++ ++ S+L   Y   EI +VDDGSS+  +  L EE+    ++
Sbjct: 90   YKVAAVIPSYNEDAESLLETLKSVLAQTYPLSEIYIVDDGSSNTDAIQLIEEY----VNR 145

Query: 65   TIDSIIATQTCKDVFQRQVGKVKLTLIVKENGGKGDALNMGINAANYDYFLCLDADSMLQ 124
             +D       C++V       V  +L+   N GK  A       ++ D FL +D+D+ +
Sbjct: 146  EVD------ICRNVI------VHRSLV---NKGKRHAQAWAFERSDADVFLTVDSDTYIY 190

Query: 125  VDSLSQISKSIQVDPTVIAVGGLVQVAQGVKIEQGKVASYRLPWRIIPCAQALEYDSSFL 184
             ++L ++ KS   D TV A               G + +     ++      + YD++F
Sbjct: 191  PNALEELLKSFN-DETVYAA-------------TGHLNARNRQTNLLTRLTDIRYDNAF- 235

Query: 185  GARIFLDYLRANLII-SGAFGLFKKD-LVKAVGGYDTQT-------LGEDMELVMKLHFF 235
            G       L  N+++ SG   +++++ ++  +  Y  QT       +G+D  L
Sbjct: 236  GVERAAQSLTGNILVCSGPLSIYRREVIIPNLERYKNQTFLGLPVSIGDDRCLT------ 289

Query: 236  CRNNNIPY-RICYETDAVCWSQAPTNLGDLRKQRRRWYLGLY-QCLKKYKSIFANYRFGA 293
              N  I   R  Y++ A C +  P  L    KQ+ RW    + + +   K I +N
Sbjct: 290  --NYAIDLGRTVYQSTARCDTDVPFQLKSYLKQQNRWNKSFFKESIISVKKILSN----P 343

Query: 294  VGSISYIYYILFELLTPFIECFGIVIIFLSLLFNQLNIPFFISLVSLYIFYCV--LITLS 351
            + ++  I+ ++  ++        +++   +LLFNQ      + L+ L+ F  +  ++ L
Sbjct: 344  IVALWTIFEVVMFMM--------LIVAIGNLLFNQ---AIQLDLIKLFAFLSIIFIVALC 392

Query: 352  SFLHRIYSQQLVIGILDIVKVFYIAVFRYLILHPVLT 388
              +H +        +  +  + ++ V + L L+ + T
Sbjct: 393  RNVHYMIKHPASFLLSPLYGILHLFVLQPLKLYSLCT 429
```

A related GBS gene <SEQ ID 8655> and protein <SEQ ID 8656> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      -5.18
GvH: Signal Score (-7.5): -4.91
     Possible site: 14
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 4 value:  -7.48 threshold:  0.0
   INTEGRAL    Likelihood = -7.48   Transmembrane   310 - 326 ( 302 - 335)
   INTEGRAL    Likelihood = -7.32   Transmembrane   362 - 378 ( 361 - 380)
   INTEGRAL    Likelihood = -7.11   Transmembrane   334 - 350 ( 329 - 355)
   INTEGRAL    Likelihood = -2.28   Transmembrane   381 - 397 ( 380 - 397)
   PERIPHERAL  Likelihood =  1.22     140
 modified ALOM score:   2.00

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3994(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00238(331 - 1401 of 1866)
GP|5813901|gb|AAD52055.1|AF086783_3|AF086783(52 - 367 of 412) IcaA {Staphylococcus aureus}
%Match = 10.3
%Identity = 34.8  %Similarity = 55.9
Matches = 109  Mismatches = 128  Conservative Sub.s = 66

150       180       210       240       270       300       330       360
VAMRRSSKLNLGVRPPFACLR**AVFNTANISSKVVR*TPTRRLNRTSVNCLLAS*FIELLYHILFRRKSIVPAYNESTT
                                                                      ::  |||| |
                   MQFFNFLLFYPVFMSIYWIVGSIYFYFTREIRYSLNKKPDINVDELEGITFLLACYNESET
                           10        20        30        40        50        60

390       420       450       471       501       531       561       591
IVSSIDSLLHLDYEAYEIIVVDDGSSDNTSDVL---KEEFALMKISNTIDSIIATQTCKDVFQRQVGKVKLTLIVKENGG
|  :: ::| | ||  |||:::|||||||::::   ||   :: :                             ::|| |
IEDTLSNVLALKYEKKEIIIINDGSSDNTAELIYKIKENNDFIFVD----------------------------LQENRG
                 80        90       100                                   110

621       651       681       711       741       771       801       831
KGDALNMGINAANYDYFLCLDADSMLQVDSLSQISKSIQVDPTVIAVGGLVQVAQGVKIEQGKVASYRLPWRIIPCAQAL
| :||| ||  |:||| :|||||:::  |:   : :: : || : || |  ::     |              :   | :
KANALNQGIKQASYDYVMCLDADTIVDQDAPYYMIENFKHDPKLGAVTGNPRIRNKSSI--------------LGKIQTI
              130       140       150       160       170

861       891       918       948       978       1008      1038      1068
EYDSSFLGARIFLDYLRANL-IISGAFGLFKKDLVKAVGGYDTQTLGEDMELVMKLHFFCRNNNIPYRICYETDAVCWSQ
|| :|::|       |   :  ||| | ||||  |  || :||  : ||: :  |||         ||| ||  |:||
EY-ASLIGCIKRSQTLAGAVNTISGVFTLFKKSAVVDVGYWDTDMITEDIAVSWKLH------LRGYRIKYEPLAMCWML
         190       200       210       220       230             240       250

1098      1128      1155                1194      1224      1254      1284
APTNLGDLRKQRRRWYLGLYQCL-KKYKSIFANYRFG-------AVGSISYIYYILFELLTPFIECFGIVIIFLSLLFNQ
 |  || | ||| ||  | :: | : : |     ||         : || ::| :|: |   ||        ||   |
VPETLGGLWKQRVRWAQGGHEVLLRDFFSTMKTKRFPLYILMFEQIISILWVYIVLLYLGYLFI-----TANFLDYTFMT
               270       280       290       300       310            320

1311      1341      1371      1401      1431      1461      1491      1521
LNIP-FFISLVSLYIFYCVLITLSSFLHRIYSQQLVIGILDIVKVFYIAVFRYLILHPVLTFVKVASVIGYKNKKMVWGH
 :   |::|  ::     :  |:: |:   | :: : |::
YSFSIFLLSSFTMTFINVIQFTVALFIDSRYEKKNMAGLIFVSWYPTVYWIINAAVVLVAFPKALKRKRGGYATWSSPDR
          340       350       360       370       380       390       400
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 733

A DNA sequence (GBSx0779) was identified in *S.agalactiae* <SEQ ID 2249> which encodes the amino acid sequence <SEQ ID 2250>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2014(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA22725 GB:AL035161 hypothetical protein SC9C7.13c
          [Streptomyces coelicolor A3(2)]
 Identities = 35/153 (22%), Positives = 64/153 (40%), Gaps = 5/153 (3%)
```

```
Query: 5   IRRARLGDEVNLAYIQTESWKAAFGKILPEDIIQKTTEIEPAITMYQQLLHKEVGKGYIL 64
           +R   L D  ++ I+   W++A+  ++P+  +        A            G+  ++
Sbjct: 10  VREMTLADCDRVSLIRVRGWQSAYRGLMPQPYLDAMDPAADAERRRSLFARPPEGRVNLV 69

Query: 65  EVDSNPHCMAWWD----KSREDGMLDYAELICIHSLKEGWGKGYGSQMMNHVLSEIQQAG 120
             D     + W      +  E    D AEL  ++     +G G G  +    +   + AG
Sbjct: 70  AEDEGGEVVGWACHGPYRDGEARTAD-AELYALYVDAARFGAGIGRALAGESVRRCRAAG 128

Query: 121 YNKVILWVFTENTRARKFYDRFGFSFKGKSKTY 153
           + +++LWV   N RAR+FYDR GF   G  + +
Sbjct: 129 HARMLLWVLKGNVRARRFYDRAGFRPDGAEEPF 161
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 734

A DNA sequence (GBSx0780) was identified in *S.agalactiae* <SEQ ID 2251> which encodes the amino acid sequence <SEQ ID 2252>. This protein is predicted to be a DNA-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1162(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 735

A DNA sequence (GBSx0781) was identified in *S.agalactiae* <SEQ ID 2253> which encodes the amino acid sequence <SEQ ID 2254>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2589(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10037> which encodes amino acid sequence <SEQ ID 10038> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2255> which encodes the amino acid sequence <SEQ ID 2256>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2767(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 80/86 (93%), Positives = 84/86 (97%)

Query: 6  LKTIKENNMTFEEILPGLKAKKKYVRTGWGGAENYVQLFDTLEVNGKVLQATPYFLINVT 65
           + +IKENNMTFEEILPGLKAKKKYVRTGWGGAENYVQLFDTLEV+GKVLQATPYFLI+VT
Sbjct: 3  ISSIKENNMTFEEILPGLKAKKKYVRTGWGGAENYVQLFDTLEVDGKVLQATPYFLIHVT 62

Query: 66 GEGEGFSMWAPTPCDVLAEDWIEVND 91
           G GEGFSMWAPTPCDVLAEDWIEVND
Sbjct: 63 GAGEGFSMWAPTPCDVLAEDWIEVND 88
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 736

A DNA sequence (GBSx0782) was identified in *S.agalactiae* <SEQ ID 2257> which encodes the amino acid sequence <SEQ ID 2258>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA85256 GB:AB021978 3-oxoacyl-[acyl carrier protein] reductase
           homolog [Moritella marina]
 Identities = 82/239 (34%), Positives = 125/239 (51%), Gaps = 15/239 (6%)

Query: 2   TKVVLVTGCASGIGYAQAQYFLKQGYQVYGVDKSDKPNLN-----GNFNF-IKLDLSSDL 55
           +K VLVTG + GIG A A++F K G  V G   S +         G+  F ++L+++S
Sbjct: 5   SKTVLVTGASRGIGRAIAEHFAKLGATVIGTATSAQGAERIGAYLGDAGFGLELNVTSQD 64

Query: 56  S------PLFTMVPTVDILCNTAGILDAYKPLLEVSDEELEHLFDINFFVTVRLTRHYLR 109
           S       + T V  +DIL N AGI  A    L + ++E  ++ D N     RL +  LR
Sbjct: 65  SVDALYAEIKTQVGHIDILVNNAGIT-ADNIFLRMKEDEWCNVIDTNLTSLYRLCKPCLR 123

Query: 110 RMVEKKSGIIINMCSIASFIAGGGGAAYTSSKHALAGFTRQLALDYAKDCIQIFGIAPGA 169
            M++++ G IIN+ S+      GG A Y ++K  L GFT+ LA + A   I +  +APG
Sbjct: 124 GMMKQRHGRIINIGSVVGTTGNGGQANYAAAKSGLLGFTKSLASEVASRGITVNAVAPGF 183

Query: 170 VQTAMTASDFEPGGLAEWVASETPIGRWTKPSEVAELTGFLASGKARSMQGEIVKIDGG 228
           ++T MTA   E     + + ++ P  R    +E+AE  GFLAS  A  + GE + ++GG
Sbjct: 184 IETDMTAELTEE--QKQTILAQVPTSRLGSTTEIAETVGFLASDGASYITGETIHVNGG 240
```

There is also homology to SEQ IDs 2628 and 7170.

A related sequence was also identified in GAS <SEQ ID 9107> which encodes the amino acid sequence <SEQ ID 9108>. Analysis of this protein sequence reveals the following:

```
     Possible site: 19
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
```

```
  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 206/232 (88%), Positives = 224/232 (95%)

Query: 1   MTKVVLVTGCASGIGYAQAQYFLKQGYQVYGVDKSDKPNLNGNFNFIKLDLSSDLSPLFT 60
           MTKVVLVTGCASGIGYAQA+YFLKQG+ VYGVDKSDKP+L+GNF+FIKLDLSS+L+PLF
Sbjct: 4   MTKVVLVTGCASGIGYAQARYFLKQGHHVYGVDKSDKPDLSGNFHFIKLDLSSELAPLFK 63

Query: 61  MVPTVDILCNTAGILDAYKPLLEVSDEELEHLFDINFFVTVRLTRHYLRRMVEKKSGIII 120
           +VP+VDILCNTAGILDAYKPLL+VSDEE+EHLFDINFF TV+LTRHYLRRMVEK+SG+II
Sbjct: 64  VVPSVDILCNTAGILDAYKPLLDVSDEEVEHLFDINFFATVKLTRHYLRRMVEKQSGVII 123

Query: 121 NMCSIASFIAGGGGAAYTSSKHALAGFTRQLALDYAKDCIQIFGIAPGAVQTAMTASDFE 180
           NMCSIASFIAGGGG AYTSSKHALAGFTRQLALDYAKD I IFGIAPGAV+TAMTA+DFE
Sbjct: 124 NMCSIASFIAGGGGVAYTSSKHALAGFTRQLALDYAKDQIHIFGIAPGAVKTAMTANDFE 183

Query: 181 PGGLAEWVASETPIGRWTKPSEVAELTGFLASGKARSMQGEIVKIDGGWSLK 232
           PGGLA+WVA ETPIGRWTKP EVAELTGFLASGKARSMQGEIVKIDGGW+LK
Sbjct: 184 PGGLADWVARETPIGRWTKPDEVAELTGFLASGKARSMQGEIVKIDGGWTLK 235
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9063> which encodes amino acid sequence <SEQ ID 9064>. An alignment of the GAS and GBS sequences follows:

```
Score = 83.1 bits (202), Expect = 4e-18
Identities = 72/258 (27%), Positives = 106/258 (40%), Gaps = 36/258 (13%)

Query: 6   EVAFITGAASGIGKQIGETLLKEGKTVVFSDINQE-----KLDQVVADYTKEGYDAFSVV 60
           +V  +TG ASGIG    +  LK+G  V   D + +     + +    D + +    F++V
Sbjct: 3   KVVLVTGCASGIGYAQAQYFLKQGYQVYGVDKSDKPNLNGNFNFIKLDLSSDLSPLFTMV 62

Query: 61  CDVTKEEAINAAIDTVVEKYGRIDILVNNAG-LQHVAMIEDFPTEKFEFMIKIMLTAPFI 119
                                 +DIL N AG L     + +   E+ E +  I
Sbjct: 63  --------------------PTVDILCNTAGILDAYKPLLEVSDEELEHLFDINFFVTVR 102

Query: 120 AIKRAFPTMKAQKHGRIINMASINGVIGFAGKSAYNSAKHGLIGLTKVTALEAADSGITV 179
             +     M  +K G IINM SI   I   G +AY S+KH L G T+  AL+ A   I +
Sbjct: 103 LTRHYLRRMVEKKSGIIINMCSIASFIAGGGGAAYTSSKHALAGFTRQLALDYAKDCIQI 162

Query: 180 NAICPGYVDTPLVRGQFEDLSKTRGIPLENVLEEVLYPLVPQKRLIDVQEIADYVSFLAS 239
             I PG V T +    FE             L E +    P  R     E+A++   FLAS
Sbjct: 163 FGIAPGAVQTAMTASDFE----------PGGLAEWVASETPIGRWTKPSEVAELTGFLAS 212

Query: 240 DKAKGVTGQACILDGGYT 257
            KA+ + G+   +DGG++
Sbjct: 213 GKARSMQGEIVKIDGGWS 230
```

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 2259> which encodes the amino acid sequence <SEQ ID 2260>. An alignment of the GAS and GBS sequences follows:

```
Score =  427 bits (1086), Expect = e-122
Identities = 206/232 (88%), Positives = 224/232 (95%)

Query: 4   MTKVVLVTGCASGIGYAQARYFLKQGHHVYGVDKSDKPDLSGNFHFIKLDLSSELAPLFK 63
           MTKVVLVTGCASGIGYAQA+YFLKQG+ VYGVDKSDKP+L+GNF+FIKLDLSS+L+PLF
Sbjct: 1   MTKVVLVTGCASGIGYAQAQYFLKQGYQVYGVDKSDKPNLNGNFNFIKLDLSSDLSPLFT 60

Query: 64  VVPSVDILCNTAGILDAYKPLLDVSDEEVEHLFDINFFATVKLTRHYLRRMVEKQSGVII 123
           +VP+VDILCNTAGILDAYKPLL+VSDEE+EHLFDINFF TV+LTRHYLRRMVEK+SG+II
Sbjct: 61  MVPTVDILCNTAGILDAYKPLLEVSDEELEHLFDINFFVTVRLTRHYLRRMVEKKSGIII 120

Query: 124 NMCSIASFIAGGGGVAYTSSKHALAGFTRQLALDYAKDQIHIFGIAPGAVKTAMTANDFE 183
           NMCSIASFIAGGGG AYTSSKHALAGFTRQLALDYAKD I IFGIAPGAV+TAMTA+DFE
```

```
Sbjct: 121 NMCSIASFIAGGGGAAYTSSKHALAGFTRQLALDYAKDCIQIFGIAPGAVQTAMTASDFE 180

Query: 184 PGGLADWVARETPIGRWTKPDEVAELTGFLASGKARSMQGEIVKIDGGWTLK 235
            PGGLA+WVA ETPIGRWTKP EVAELTGFLASGKARSMQGEIVKIDGGW+LK
Sbjct: 181 PGGLAEWVASETPIGRWTKPSEVAELTGFLASGKARSMQGEIVKIDGGWSLK 232
```

SEQ ID 2258 (GBS251) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 2; MW 21.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 6; MW 52kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 737

A DNA sequence (GBSx0783) was identified in *S.agalactiae* <SEQ ID 2261> which encodes the amino acid sequence <SEQ ID 2262>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.82   Transmembrane   62 -  78 (  62 -  79)

----- Final Results -----
              bacterial membrane --- Certainty=0.2529(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 738

A DNA sequence (GBSx0784) was identified in *S.agalactiae* <SEQ ID 2263> which encodes the amino acid sequence <SEQ ID 2264>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1495(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA20397 GB:AL031317 SC6G4.19c, unknown, len: 190 aa; contains
           Pro-Ser- r ich domain at N-terminus [Streptomyces
           coelicolor A3(2)]
 Identities = 26/80 (32%), Positives = 44/80 (54%), Gaps = 5/80 (6%)

Query: 1   MDSNDEAICIIEITKVDIVPFKDVSADHAFKEGEGDKTLEWWRKAHIDFF-----KPYFE 55
            +DS +  + +IE+T+V +VP  +V   HA  EGEGD ++  WR  H  F+      +
Sbjct: 103 VDSRERPVAVIEVTEVRVVPLAEVDLAHAVDEGEGDTSVAGWRAGHERFWHGAEMRAALG 162

Query: 56  EFGLMFSEDSRIVLEEFQVV 75
            + G    + + +VLE F++V
Sbjct: 163 DPGFTVDDATPVVLERFRIV 182
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 739

A DNA sequence (GBSx0785) was identified in *S.agalactiae* <SEQ ID 2265> which encodes the amino acid sequence <SEQ ID 2266>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.49   Transmembrane    3 -  19 (   3 -  19)

----- Final Results -----
               bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06422 GB:AP001516 unknown conserved protein [Bacillus halodurans]
 Identities = 133/315 (42%), Positives = 191/315 (60%), Gaps = 4/315 (1%)

Query: 1    MKLAVLGTGMIVKEVLPVLQKIEGIDLVAILSTVRSLETAKDLAKEYNMSLATSEYKAVL 60
            MK+A +GTG IV+  L  L  I+G   VA+ S  R   TAK LA +YN+    + +  +L
Sbjct: 1    MKIATVGTGPIVEAFLSALDDIDGPMCVAMYS--RKETTAKPLADQYNIPTIYTHFDHML 58

Query: 61   DNEEIDTVYIGLPNHLHFDYAKEALLAGKHVICEKPFTLEASQLEELVSIANTRQLILLE 120
             +  ++ VY+   PN LH+ +A +AL   KHVICEKPFT  A +LE L+S+A   +L+L E
Sbjct: 59   ADPNVEVVYVASPNSLHYQHALQALEHRKHVICEKPFTSTARELEHLISVARKNELMLFE 118

Query: 121  AITNQYLPNFDLVKEHLSNLGDIKIVECNYSQYSSRYDAFKRGEIAPAFNPEMGGGALRD 180
            AIT  +LPN+ L+KE++  LG IK+++CNYSQYSSRYD F  GE    FNP  GGAL D
Sbjct: 119  AITTIHLPNYQLIKENIHKLGSIKMIQCNYSQYSSRYDRFLSGETPNVFNPAFSGGALMD 178

Query: 181  LNIYNLHLVIGLFGEPITAQYLPNIE-RGIDTSGVLVLDYGHFKTVCIGAKDCSAEVKST 239
            +N+YN+H V+ LFG P  A Y+ N    GIDTSGVLVL Y HF + C+G KD  +
Sbjct: 179  INVYNIHFVMNLFGPPEAAHYIANQHANGIDTSGVLVLKYPHFISECVGCKDTQSMNFVL 238

Query: 240  IQGDKGSIAILGPTNTMPKISLTMNGQESHVYQLNGDRHRMHDEFVIFEGIISNLDFKRA 299
            IQG+KG I +    N    + + ++ Q S +     D    ++    +E +    +F++
Sbjct: 239  IQGEKGYIHVENGANGCRNVKIYLDDQTSELNAQTNDNLLYYETRTFYE-MYQAKNFEKC 297

Query: 300  AQALEHSRTVMKVLD 314
             + L +S +VM+V++
Sbjct: 298  YELLSYSHSVMRVME 312
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 719> which encodes the amino acid sequence <SEQ ID 720>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 233/314 (74%), Positives = 269/314 (85%)

Query: 1    MKLAVLGTGMIVKEVLPVLQKIEGIDLVAILSTVRSLETAKDLAKEYNMSLATSEYKAVL 60
            MKLAVLGTGMIVKEVLPVLQKI+GIDLVAILSTVRSL TAKDLAK ++M LATS+Y+A+L
```

```
Sbjct: 1   MKLAVLGTGMIVKEVLPVLQKIDGIDLVAILSTVRSLTTAKDLAKAHHMPLATSKYEAIL 60

Query: 61  DNEEIDTVYIGLPNHLHFDYAKEALLAGKHVICEKPFTLEASQLEELVSIANTRQLILLE 120
            NEEIDTVYIGLPNHLHF YAKEALLAGKHVICEKPFT+ A +L+ELV IA  R+LILLE
Sbjct: 61  GNEEIDTVYIGLPNHLHFAYAKEALLAGKHVICEKPFTMTAGELDELVVIARKRKLILLE 120

Query: 121 AITNQYLPNFDLVKEHLSNLGDIKIVECNYSQYSSRYDAFKRGEIAPAFNPEMGGGALRD 180
           AITNQYL N   +KEHL  LGDIKIVECNYSQYSSRYDAFKRG+IAPAFNP+MGGGALRD
Sbjct: 121 AITNQYLSNMTFIKEHLDQLGDIKIVECNYSQYSSRYDAFKRGDIAPAFNPKMGGGALRD 180

Query: 181 LNIYNLHLVIGLFGEPITAQYLPNIERGIDTSGVLVLDYGHFKTVCIGAKDCSAEVKSTI 240
           LNIYN+H V+GLFG P T QYL N+E+GIDTSG+LV+DY  FK VCIGAKDC+AE+KSTI
Sbjct: 181 LNIYNIHFVVGLFGRPKTVQYLANVEKGIDTSGMLVMDYEQFKVVCIGAKDCTAEIKSTI 240

Query: 241 QGDKGSIAILGPTNTMPKISLTMNGQESHVYQLNGDRHRMHDEFVIFEGIISNLDFKRAA 300
           QG+KGS+A+LG TNT+P++ L+++G E  V   N   HRM++EFV F  +I   DF++
Sbjct: 241 QGNKGSLAVLGATNTLPQVQLSLHGHEPQVINHNKHDHRMYEEFVAFRDMIDQRDFEKVN 300

Query: 301 QALEHSRTVMKVLD 314
           QALEHSR VM VL+
Sbjct: 301 QALEHSRAVMAVLE 314
```

SEQ ID 2266 (GBS342) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 10; MW 36.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 2; MW 61kDa).

GBS342-GST was purified as shown in Figure 226, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 740

A DNA sequence (GBSx0786) was identified in *S.agalactiae* <SEQ ID 2267> which encodes the amino acid sequence <SEQ ID 2268>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0499(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12535 GB:Z99107 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 41/127 (32%), Positives = 63/127 (49%), Gaps = 11/127 (8%)

Query: 1   MISSIGQVMLYVSNVEASADFWKNKVGFERVEKQTQGDYVTYI-VAPKLDSEVSFVLHDK 59
           MI  IG V +YV + + +  FW  KVGF+       G   +++ VAPK  +E   V++ K
Sbjct: 1   MIKQIGTVAVYVEDQQKAKQFWTEKVGFDIAADHPMGPEASWLEVAPK-GAETRLVIYPK 59

Query: 60  AIIAQMSPELDLATPSILFETTDIDSTYQELTAN--EVMTNP-IVDMGSMRVFNFSDNDN 116
           A    M    +    SI+FE  DI  TY+++  N  E +  P  ++ G+     F D D
Sbjct: 60  A----MMKGSEQMKASIVFECEDIFGTYEKMKTNGVEFLGEPNQMEWGTF--VQFKDEDG 113

Query: 117 NYFAIRE 123
           N F ++E
Sbjct: 114 NVFLLKE 120
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 741

A DNA sequence (GBSx0787) was identified in *S.agalactiae* <SEQ ID 2269> which encodes the amino acid sequence <SEQ ID 2270>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3402(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04569 GB:AP001510 unknown conserved protein in others
          [Bacillus halodurans]
 Identities = 46/144 (31%), Positives = 83/144 (56%), Gaps = 10/144 (6%)

Query: 1   MVKALETYIVTNGNGRQAVDFYKDVFQADLVNMMTWEEM--DPNC--LEDRKDLIINAQL 56
           M+  +  Y++ +G+G+ A++FY+D   A+++ + T+ ++   PN       KDLI++A L
Sbjct: 1   MILTMNPYLMLDGDGQAAIEFYQDALNAEVITIQTYGDLPEQPNSPMASVNKDLILHAHL 60

Query: 57  IFDGIRLQISDENPD-----FVYQAGKNVTAAIIVGSVEEAREIYEKLKKSAQEVQLELQ 111
               + L ISD+  D     F   +G  VT A+   +VE   E+++KL    +E+   L+
Sbjct: 61  KLGEMDLMISDQCLDVDPERFPQHSGSPVTIALTTNNVEMTTEVFQKLASGGEEIA-PLE 119

Query: 112 ETFWSPAYANLVDQFGVMWQISTE 135
           +TF+SP Y  + D+FG+ W +ST+
Sbjct: 120 KTFFSPLYGQVTDKFGITWHVSTQ 143
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 742

A DNA sequence (GBSx0788) was identified in *S.agalactiae* <SEQ ID 2271> which encodes the amino acid sequence <SEQ ID 2272>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB03784 GB:AP001507 UDP-N-acetylglucosamine pyrophosphorylase
          [Bacillus halodurans]
 Identities = 238/453 (52%), Positives = 322/453 (70%), Gaps = 1/453 (0%)

Query: 1   MSN-YAIILAAGKGTRMKSDLPKVMHKVSGITMLEHVFRSVQAIEPSKIVTVIGHKAELV 59
           MSN +A+ILAAG+GTRMKS L KV+H V G  M++HV   V A+   +IVT+IGH A+ V
Sbjct: 1   MSNRFAVILAAGQGTRMKSKLYKVLHSVCGKPMVQHVVDQVSALGFDEIVTIIGHGADAV 60

Query: 60  RDVLGDKSEFVMQTEQLGTGHAVMMAEEELATSKGHTLVIAGDTPLITGESLKNLIDFHV 119
           +  LG++  + +Q EQLGTGHAV+ AE  L   +G T+V+ GDTPL+T E++ +++ +H
```

```
Sbjct: 61  KSQLGERVSYALQEEQLGTGHAVLQAESALGGRRGVTIVLCGDTPLLTAETIDHVMSYHE 120

Query: 120 NHKNVATILTADAANPFGYGRIIRNSDDEVTKIVEQKDANDFEQQVKEINTGTYVFDNQS 179
              +  AT+LTA+ A+P GYGRI+RN    V +IVE KDA   E+Q+ E+NTGTY FDN++
Sbjct: 121 EEQAKATVLTAELADPTGYGRIVRNDKGLVERIVEHKDATSEEKQITEVNTGTYCFDNEA 180

Query: 180 LFEALKDINTNNAQGEYYLTDVIGIFKEAGKKVGAYKLRDFDESLGVNDRVALATAEKVM 239
           LF+ALK++  NNAQGEYYL DVI I +  G+KV AYK    +E+LGVNDRVALA AE+VM
Sbjct: 181 LFQALKEVGNNNAQGEYYLPDVIQILQTKGEKVAAYKTAHVEETLGVNDRVALAQAEQVM 240

Query: 240 RHRIARQHMVNGVTVVNPDSAYIDIDVEIGEESVIEPNVTLKGQTKIGKGTLLTNGSYLV 299
           + RI    M  GVT ++P+  Y+  D  IG+++VI P    + GQT IG+G +L   + L
Sbjct: 241 KRRINEAWMRKGVTFIDPEQTYVSPDATIGQDTVIYPGTMVLGQTTIGEGCVLGPHTELK 300

Query: 300 DAQVGNDVTITNSMVEESIISDGVTVGPYAHIRPGTSLAKGVHIGNFVEVKGSQIGENTK 359
           D+++GN   +  S+V  S + + V++GP++HIRP + +   V IGNFVEVK S IG+ +K
Sbjct: 301 DSKIGNKTAVKQSVVHNSEVGERVSIGPFSHIRPASMIHDDVRIGNFVEVKKSTIGKESK 360

Query: 360 AGHLTYIGNAEVGCDVNFGAGTITVNYDGQNKFKTEIGSNVFIGSNSTLIAPLEIGDNAL 419
           A HL+YIG+AEVG  VNF  G+ITVNYDG+NKF T+I  + FIG NS LIAP+ IG  AL
Sbjct: 361 ASHLSYIGDAEVGERVNFSCGSITVNYDGKNKFLTKIEDDAFIGCNSNLIAPVTIGKGAL 420

Query: 420 TAAGSTITDNVPIDSIAIGRGRQVNKEGYANKK 452
            AAGSTIT++VP D+++I R RQ NKE Y  KK
Sbjct: 421 IAAGSTITEDVPSDALSIARARQTNKEHYVTKK 453
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2273> which encodes the amino acid sequence <SEQ ID 2274>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0461(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 345/458 (75%), Positives = 398/458 (86%)

Query: 1   MSNYAIILAAGKGTRMKSDLPKVMHKVSGITMLEHVFRSVQAIEPSKIVTVIGHKAELVR 60
           M+NYAIILAAGKGTRM SDLPKV+HKVSG+TMLEHVFRSV+AI P K VTVIGHK+E+VR
Sbjct: 1   MTNYAIILAAGKGTRMTSDLPKVLHKVSGLTMLEHVFRSVKAISPEKSVTVIGHKSEMVR 60

Query: 61  DVLGDKSEFVMQTEQLGTGHAVMMAEEELATSKGHTLVIAGDTPLITGESLKNLIDFHVN 120
            VL D+S FV QTEQLGTGHAVMMAE +L   +GHTLVIAGDTPLITGESLK+LIDFHVN
Sbjct: 61  AVLADQSAFVHQTEQLGTGHAVMMAETQLEGLEGHTLVIAGDTPLITGESLKSLIDFHVN 120

Query: 121 HKNVATILTADAANPFGYGRIIRNSDDEVTKIVEQKDANDFEQQVKEINTGTYVFDNQSL 180
           HKNVATILTA A +PFGYGRI+RN D EV KIVEQKDAN++EQQ+KEINTGTYVFDN+ L
Sbjct: 121 HKNVATILTATAQDPFGYGRIVRNKDGEVIKIVEQKDANEYEQQLKEINTGTYVFDNKRL 180

Query: 181 FEALKDINTNNAQGEYYLTDVIGIFKEAGKKVGAYKLRDFDESLGVNDRVALATAEKVMR 240
           FEALK I TNNAQGEYYLTDV+ IF+   +KVGAY LRDF+ESLGVNDRVALA AE VMR
Sbjct: 181 FEALKCITTNNAQGEYYLTDVVAIFRANKEKVGAYILRDFNESLGVNDRVALAIAETVMR 240

Query: 241 HRIARQHMVNGVTVVNPDSAYIDIDVEIGEESVIEPNVTLKGQTKIGKGTLLTNGSYLVD 300
            RI ++HMVNGVT  NP++ YI+ DVEI  + +IE NVTLKG+T IG GT+LTNG+Y+VD
Sbjct: 241 QRITQKHMVNGVTFQNPETVYIESDVEIAPDVLIEGNVTLKGRTHIGSGTVLTNGTYIVD 300

Query: 301 AQVGNDVTITNSMVEESIISDGVTVGPYAHIRPGTSLAKGVHIGNFVEVKGSQIGENTKA 360
           +++G++  +TNSM+E S+++ GVTVGPYAH+RPGT+L + VHIGNFVEVKGS IGE TKA
Sbjct: 301 SEIGDNCVVTNSMIESSVLAAGVTVGPYAHLRPGTTLDREVHIGNFVEVKGSHIGEKTKA 360

Query: 361 GHLTYIGNAEVGCDVNFGAGTITVNYDGQNKFKTEIGSNVFIGSNSTLIAPLEIGDNALT 420
           GHLTYIGNA+VG  VN GAGTITVNYDGQNK++T IG + FIGSNSTLIAPLE+GD+ALT
```

```
Sbjct: 361 GHLTYIGNAQVGSSVNVGAGTITVNYDGQNKYETVIGDHAFIGSNSTLIAPLEVGDHALT 420

Query: 421 AAGSTITDNVPIDSIAIGRGRQVNKEGYANKKPHHPSQ 458
           AAGSTI+  VPIDSIAIGR RQV KEGYA +  HHPS+
Sbjct: 421 AAGSTISKTVPIDSIAIGRSRQVTKEGYAKRLAHHPSR 458
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 743

A DNA sequence (GBSx0790) was identified in *S.agalactiae* <SEQ ID 2275> which encodes the amino acid sequence <SEQ ID 2276>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1366(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14293 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 92/177 (51%), Positives = 124/177 (69%), Gaps = 4/177 (2%)

Query: 4   EEKTINRQTVFDGQIIKVAVDDVELPNGLGQSKRELVFHGGAVATLAVTPEHKIVLVKQY 63
           EEKTI ++ +F G++I + V+DVELPNG   SKRE+V H GAVA LAVT E KI++VKQ+
Sbjct: 5   EEKTIAKEQIFSGKVIDLYVEDVELPNGKA-SKREIVKHPGAVAVLAVTDEGKIIMVKQF 63

Query: 64  RKAIEGISYEIPAGKLETGESGSKEEAALRELEEETGYTG-NLEILYSFYTAIGFCNEKI 122
           RK +E    EIPAGKLE GE    E  ALRELEEETGYT   L  + +FYT+ GF +E +
Sbjct: 64  RKPLERTIVEIPAGKLEKGE--EPEYTALRELEEETGYTAKKLTKITAFYTSPGFADEIV 121

Query: 123 VLYLATDLQKVENPRPQDDDEVLELLELSYEDCMQMVEKGMIQDAKTIIALQYYGLK 179
            ++LA +L  +E  R  D+DE +E++E++ ED +++VE   + DAKT  A+QY  LK
Sbjct: 122 HVFLAEELSVLEEKRELDEDEFVEVMEVTLEDALKLVESREVYDAKTAYAIQYLQLK 178
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2277> which encodes the amino acid sequence <SEQ ID 2278>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1120(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 136/182 (74%), Positives = 153/182 (83%)

Query: 1   MDFEEKTINRQTVFDGQIIKVAVDDVELPNGLGQSKRELVFHGGAVATLAVTPEHKIVLV 60
           M FEEKT+ RQTVFDG I KV VDDVELPN LGQSKREL+FH GAVA LA+TPE KIVLV
Sbjct: 1   MKFEEKTLKRQTVFDGHIFKVVVDDVELPNNLGQSKRELIFHRGAVAVLAITPERKIVLV 60

Query: 61  KQYRKAIEGISYEIPAGKLETGESGSKEEAALRELEEETGYTGNLEILYSFYTAIGFCNE 120
           KQYRKAIE +SYEIPAGKLE GE GSK +AA RELEEET YTG L  LY FYTAIGFCNE
Sbjct: 61  KQYRKAIERVSYEIPAGKLEIGEEGSKLKAAARELEEETAYTGTLTFLYEFYTAIGFCNE 120

Query: 121 KIVLYLATDLQKVENPRPQDDDEVLELLELSYEDCMQMVEKGMIQDAKTIIALQYYGLKM 180
           KI L+LATDL +V NP+PQDDDEV+E+LEL+Y++CM +V +G + DAKT+IALQYY L
Sbjct: 121 KITLFLATDLIQVANPKPQDDDEVIEVLELTYQECMDLVAQGKLADAKTLIALQYYALHF 180
```

```
Query: 181 GG 182
           GG
Sbjct: 181 GG 182
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 744

A DNA sequence (GBSx0791) was identified in *S.agalactiae* <SEQ ID 2279> which encodes the amino acid sequence <SEQ ID 2280>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-15.44   Transmembrane   70 -  86 (  64 -  88)

----- Final Results -----
               bacterial membrane --- Certainty=0.7177(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2281> which encodes the amino acid sequence <SEQ ID 2282>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-15.60   Transmembrane   65 -  81 (  58 -  83)

----- Final Results -----
               bacterial membrane --- Certainty=0.7241(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 39/89 (43%), Positives = 61/89 (67%), Gaps = 6/89 (6%)

Query: 1  MGKPLLTDDMIERSNRGEKVSGQTILDQETKIISTEDGMEQLTDENGKHIYKSRRIENAK 60
          MG+PLLTDD+IE++ R E       ++ +TK+++  +      ++    IYKSRRIENAK
Sbjct: 2  MGRPLLTDDIIEKARRMETFEPDDAVNFDTKVMTLPE------KDDKARIYKSRRIENAK 55

Query: 61 RNEFQRKLNLVLFILLILLALLFYAIFKL 89
          R++ Q KLN++L  +++L+A+L YAIF L
Sbjct: 56 RSQLQSKLNVILIAVMLLIAILVYAIFYL 84
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 745

A DNA sequence (GBSx0792) was identified in *S.agalactiae* <SEQ ID 2283> which encodes the amino acid sequence <SEQ ID 2284>. This protein is predicted to be pfs protein (pfs). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane   56 -  72 (  56 -  72)

----- Final Results -----
```

```
       bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC22869 GB:U32801 pfs protein (pfs) [Haemophilus influenzae Rd]
 Identities = 100/229 (43%), Positives = 144/229 (62%)

Query: 1   MKIGIIAAMEEELKLLVENLEDKSQETVLSNVYYSGRYGEHELVLVQSGVGKVMSAMSVA 60
           MKIGI+ AM +E+++L   + D+++  V S V + G+    ++ L+QSG+GKV +A+
Sbjct: 1   MKIGIVGAMAQEVEILKNLMADRTETRVASAVIFEGKINGKDVALLQSGIGKVAAAIGTT 60

Query: 61  ILVESFKVDAIINTGSAGAVATGLNVGDVVVADTLVYHDVDLTAFGYDYGQMSMQPLYFH 120
            L++  K D +INTGSAG VA GL VGD+V++D   YHD D+TAFGY+ GQ+   P  F
Sbjct: 61  ALLQLAKPDCVINTGSAGGVAKGLKVGDIVISDETRYHDADVTAFGYEKGQLPANPAAFL 120

Query: 121 SDKTFVSTFEAVLSKEEMISKVGLIATGDSFIAGQEKIDVIKGHFPQVLAVEMEGAAIAQ 180
           SDK      + +  K+    K GLI +GDSFI  ++KI  IK  FP V  VEME  AIAQ
Sbjct: 121 SDKKLADLAQEIAEKQGQSVKRGLICSGDSFINSEDKIAQIKADFPNVTGVEMEATAIAQ 180

Query: 181 AAQATGKPFVVVRAMSDTAAHDANITFDEFIIEAGKRSAQVLMAFLKAL 229
              A   PFVVVRA+SD     A+++F+EF+  A K+S+ +++  +  L
Sbjct: 181 VCYAFNVPFVVVRAISDGGDGKASMSFEEFLPLAAKQSSALVLGMIDRL 229
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2285> which encodes the amino acid sequence <SEQ ID 2286>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1245(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 169/229 (73%), Positives = 189/229 (81%)

Query: 1   MKIGIIAAMEEELKLLVENLEDKSQETVLSNVYYSGRYGEHELVLVQSGVGKVMSAMSVA 60
           MKIGIIAAMEEEL LL+ NL D  +  VLS  YY+GR+G+HEL+LVQSGVGKVMSAM+VA
Sbjct: 1   MKIGIIAAMEEELSLLLANLLDAQEHQVLSKTYYTGRFGKHELILVQSGVGKVMSAMTVA 60

Query: 61  ILVESFKVDAIINTGSAGAVATGLNVGDVVVADTLVYHDVDLTAFGYDYGQMSMQPLYFH 120
           ILVE FK  AIINTGSAGAVA+ L +GDVVVAD LVYHDVD TAFGY YGQM+ QPLY+
Sbjct: 61  ILVEHFKAQAIINTGSAGAVASHLAIGDVVVADRLVYHDVDATAFGYAYGQMAGQPLYYD 120

Query: 121 SDKTFVSTFEAVLSKEEMISKVGLIATGDSFIAGQEKIDVIKGHFPQVLAVEMEGAAIAQ 180
            D  FV+ F+ VL  E+   +VGLIATGDSF+AGQ+KID IK  F  VLAVEMEGAAIAQ
Sbjct: 121 CDPQFVAIFKQVLKHEKTNGQVGLIATGDSFVAGQDKIDQIKTAFSDVLAVEMEGAAIAQ 180

Query: 181 AAQATGKPFVVVRAMSDTAAHDANITFDEFIIEAGKRSAQVLMAFLKAL 229
           AA   GKPF+VVRAMSDTAAHDANITFD+FIIEAGKRSAQ LM FL+ L
Sbjct: 181 AAHTAGKPFIVVRAMSDTAAHDANITFDQFIIEAGKRSAQTLMTFLENL 229
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 746

A DNA sequence (GBSx0793) was identified in *S.agalactiae* <SEQ ID 2287> which encodes the amino acid sequence <SEQ ID 2288>. This protein is predicted to be SloR. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3777(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9405> which encodes amino acid sequence <SEQ ID 9406> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF81675 GB:AF232688 SloR [Streptococcus mutans]
 Identities = 97/175 (55%), Positives = 134/175 (76%)

Query: 1   MSEMIKKMISEQLIVKDKDLGYYLTKQGLLVVSDLYRKHRLVEVFLVNHLHYTADDIHEE 60
            +SEM+KK++ E L++KDK  GY LTK+G ++ S LYRKHRL+EVFL+NHL+YTAD+IHEE
Sbjct: 38  VSEMVKKLLLEDLVLKDKQAGYLLTKKGQILASSLYRKHRLIEVFLMNHLNYTADEIHEE 97

Query: 61  AEVLEHTVSTTFVDQLEKLLDFPQFCPHGGTIPKKGEFLVEINQMTLDQISQLGTYVISR 120
            AEVLEHTVS  FV++L+K L++P+ CPHGGTIP+ G+ LVE  + TL  ++++G Y++ R
Sbjct: 98  AEVLEHTVSDVFVERLDKFLNYPKVCPHGGTIPQHGQPLVERYRTTLKGVTEMGVYLLKR 157

Query: 121 VHDDFQLLKYLEQHRLHINDTIELTQIDPYAKTYHITYNDENLTIPERIASQIYV 175
            V D+FQLLKY+EQH L I D + L + D +A  Y I  + E L +   +ASQIY+
Sbjct: 158 VQDNFQLLKYMEQHHLKIGDELRLLEYDAFAGAYTIEKDGEQLQVTSAVASQIYI 212
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2289> which encodes the amino acid sequence <SEQ ID 2290>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2910(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 44/75 (58%), Positives = 59/75 (78%)

Query: 1   MSEMIKKMISEQLIVKDKDLGYYLTKQGLLVVSDLYRKHRLVEVFLVNHLHYTADDIHEE 60
            +SEMIKKMIS+  IVKDK  GY L  +G  +V++LYRK RL+EVFL++ L Y   ++H+E
Sbjct: 38  VSEMIKKMISQGWIVKDKAKGYLLKDKGYALVANLYRKLRLIEVFLIHQLGYNTQEVHQE 97

Query: 61  AEVLEHTVSTTFVDQ 75
            AEVLEHTVS +F+D+
Sbjct: 98  AEVLEHTVSDSFIDR 112
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 747

A DNA sequence (GBSx0794) was identified in *S.agalactiae* <SEQ ID 2291> which encodes the amino acid sequence <SEQ ID 2292>. This protein is predicted to be undecaprenyl pyrophosphate synthetase (uppS). Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3569(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9435> which encodes amino acid sequence <SEQ ID 9436> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13526 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 88/165 (53%), Positives = 118/165 (71%), Gaps = 4/165 (2%)

Query: 1   MNLPVKFFDKYVPELDKNNVRVQVIGDTHKLPKATYDAMQRACLRTKHNSGLVLNFALNY 60
           M LP +F + Y+PEL + NV+V++IGD   LP  T  A+++A   T  N G++LNFALNY
Sbjct: 100 MKLPEEFLNTYLPELVEENVQVRIIGDETALPAHTLRAIEKAVQDTAQNDGMILNFALNY 159

Query: 61  GGRSEITNAIKEIAQDVLEAKLNPDDITEDLVANHLMTNSLPYLYRDPDLIIRTSGELRL 120
           GGR+EI +A K +A+ V E  LN +DI E L + +LMT SL    +DP+L+IRTSGE+RL
Sbjct: 160 GGRTEIVSAAKSLAEKVKEGSLNIEDIDESLFSTYLMTESL----QDPELLIRTSGEIRL 215

Query: 121 SNFLPWQSAYSEFYFTPVLWPDFKKDELHKAIVDYNQRHRRFGSV 165
           SNF+ WQ AYSEF FT VLWPDFK+D   +A+ ++ QR RRFG +
Sbjct: 216 SNFMLWQVAYSEFVFTDVLWPDFKEDHFLQALGEFQQRGRRFGGI 260
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2293> which encodes the amino acid sequence <SEQ ID 2294>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2073(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 125/165 (75%), Positives = 145/165 (87%)

Query: 1   MNLPVKFFDKYVPELDKNNVRVQVIGDTHKLPKATYDAMQRACLRTKHNSGLVLNFALNY 60
           MNLPV FFDKYVP L +NNV++Q+IG+T +LP+ T  A+  A  +TK N+GL+LNFALNY
Sbjct: 85  MNLPVTFFDKYVPVLHENNVKIQMIGETSRLPEDTLAALNAAIDKTKRNTGLILNFALNY 144

Query: 61  GGRSEITNAIKEIAQDVLEAKLNPDDITEDLVANHLMTNSLPYLYRDPDLIIRTSGELRL 120
           GGR+EIT+A++ IAQDVL+AKLNP DITEDL+AN+LMT+ LPYLYRDPDLIIRTSGELRL
Sbjct: 145 GGRAEITSAVRFIAQDVLDAKLNPGDITEDLIANYLMTDHLPYLYRDPDLIIRTSGELRL 204

Query: 121 SNFLPWQSAYSEFYFTPVLWPDFKKDELHKAIVDYNQRHRRFGSV 165
           SNFLPWQSAYSEFYFTPVLWPDFKK EL KAI DYN+R RRFG V
Sbjct: 205 SNFLPWQSAYSEFYFTPVLWPDFKKAELLKAIADYNRRQRRFGKV 249
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 748

A DNA sequence (GBSx0795) was identified in *S.agalactiae* <SEQ ID 2295> which encodes the amino acid sequence <SEQ ID 2296>. This protein is predicted to be phosphatidate cytidylyltransferase (cdsA). Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
```

```
   INTEGRAL    Likelihood = -8.65    Transmembrane  201 - 217  ( 194 - 222)
   INTEGRAL    Likelihood = -7.96    Transmembrane  175 - 191  ( 170 - 197)
   INTEGRAL    Likelihood = -5.89    Transmembrane   81 -  97  (  74 -  99)
   INTEGRAL    Likelihood = -3.03    Transmembrane   26 -  42  (  23 -  42)
   INTEGRAL    Likelihood = -2.92    Transmembrane  136 - 152  ( 135 - 153)
   INTEGRAL    Likelihood = -2.02    Transmembrane   49 -  65  (  47 -  66)
   INTEGRAL    Likelihood = -0.64    Transmembrane  248 - 264  ( 248 - 264)

----- Final Results -----
                bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06141 GB:AP001515 phosphatidate cytidylyltransferase
          [Bacillus halodurans]
 Identities = 116/266 (43%), Positives = 172/266 (64%), Gaps = 6/266 (2%)

Query: 1   MKERVIWGAVALAIFIPFLVMGGLPFQFLVGLLAMIGVSELLRMRRLEIFSFEGALAMIG 60
           MK+RV+   +   +F+ F+V+GGLPF   + ++A I +SELL+M+++  FS  GA +++
Sbjct: 1   MKQRVVTAIIFGLVFLTFVVVGGLPFTMFIIVVATIAMSELLKMKKIAPFSPMGAFSLLP 60

Query: 61  AFVLTVPLDSYLSFLPVDASLSAYGIVIFMILAGTVLNSNSYSFEDAAFPIASSFYVGIG 120
            ++L +P D +   +P    +  +   I  +L  TVL  N+++F++A F I SS Y+G G
Sbjct: 61  MWMLLLPNDWFKVVIPDFTKVEIFIFFILFLLLLTVLTKNTFTFDEAGFVILSSAYIGYG 120

Query: 121 FQNLVSARMA---GIDKVLLALFIVWATDIGAYMIGRQFGQRKLLPSVSPNKTIEGSLGG 177
           F  L+ +R     G+  V   LF++WATD GAY  GR FG+ KL P +SPNKTIEGS+GG
Sbjct: 121 FHFLLLSREIPEIGLPLVFFVLFVIWATDSGAYFAGRAFGKHKLWPHISPNKTIEGSIGG 180

Query: 178 IASAIVVAFFFMLFDKTVYAPHSFLVMLVLVAIFSIFGQFGDLVESSIKRHFGVKDSGKL 237
           I  A+++   F           S+ V L ++ + S+FGQ GDLVES++KRH+ VKDSG +
Sbjct: 181 IILAVIIGSLFYWIMPLF---SSYGVALAVIVVASVFGQLGDLVESALKRHYAVKDSGTV 237

Query: 238 IPGHGGILDRFDSMIFVFPIMHFFGL 263
           +PGHGGILDRFDS+I+V PI+H   L
Sbjct: 238 LPGHGGILDRFDSLIYVMPILHLLHL 263
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2297> which encodes the amino acid sequence <SEQ ID 2298>. Analysis of this protein sequence reveals the following:

```
     Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.98    Transmembrane  175 - 191  ( 170 - 197)
   INTEGRAL    Likelihood = -8.97    Transmembrane    5 -  21  (   4 -  42)
   INTEGRAL    Likelihood = -6.85    Transmembrane  201 - 217  ( 197 - 222)
   INTEGRAL    Likelihood = -6.53    Transmembrane   81 -  97  (  79 -  99)
   INTEGRAL    Likelihood = -4.73    Transmembrane   49 -  65  (  47 -  71)
   INTEGRAL    Likelihood = -3.40    Transmembrane  136 - 152  ( 135 - 153)
   INTEGRAL    Likelihood = -3.24    Transmembrane   26 -  42  (  22 -  42)
   INTEGRAL    Likelihood = -1.17    Transmembrane  248 - 264  ( 248 - 264)

----- Final Results -----
                bacterial membrane --- Certainty=0.4991(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB06141 GB:AP001515 phosphatidate cytidylyltransferase
          [Bacillus halodurans]
 Identities = 125/266 (46%), Positives = 177/266 (65%), Gaps = 6/266 (2%)

Query: 1   MKERVVWGGVAVAIFLPFLIIGNLPFQLFVGVLAMIGVSELLKMKRLEVFSFEGVFAMLA 60
           MK+RVV   +   +FL F+++G LPF +F+ V+A I +SELLKMK++  FS  G F++L
Sbjct: 1   MKQRVVTAIIFGLVFLTFVVVGGLPFTMFIIVVATIAMSELLKMKKIAPFSPMGAFSLLP 60
```

```
Query: 61  AFVLAVPMDHYLTFLPIDANVAFYSLMVFFILAGTVLNSRAYSFDDAAFPIATSFYVGIG 120
            ++L +P D +   +P    V  +   + F+L  TVL    ++FD+A F I +S Y+G G
Sbjct: 61  MWMLLLPNDWFKVVIPDFTKVEIFIFFILFLLLLTVLTKNTFTFDEAGFVILSSAYIGYG 120

Query: 121 FQHLINAR---LSGIDKVFLALFIVWATDIGAYLIGRQFGRRKLLPTVSPNKTIEGSLGG 177
           F  L+ +R     G+  VF  LF++WATD GAY  GR FG+ KL P +SPNKTIEGS+GG
Sbjct: 121 FHFLLLSREIPEIGLPLVFFVLFVIWATDSGAYFAGRAFGKHKLWPHISPNKTIEGSIGG 180

Query: 178 IACAVLVSFIFMVIDRSVYAPHHFLTMLVLVALFSIFAQFGDLVESALKRHFGVKDSGKL 237
           I  AV++  +F  I   +++ +     +++VA  S+F Q GDLVESALKRH+ VKDSG +
Sbjct: 181 IILAVIIGSLFYWI-MPLFSSYGVALAVIVVA--SVFGQLGDLVESALKRHYAVKDSGTV 237

Query: 238 IPGHGGILDRFDSMIFVFPIMHLFGL 263
           +PGHGGILDRFDS+I+V PI+HL  L
Sbjct: 238 LPGHGGILDRFDSLIYVMPILHLLHL 263
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 204/264 (77%), Positives = 243/264 (91%)

Query: 1   MKERVIWGAVALAIFIPFLVMGGLPFQFLVGLLAMIGVSELLRMRRLEIFSFEGALAMIG 60
           MKERV+WG VA+AIF+PFL++G LPFQ  VG+LAMIGVSELL+M+RLE+FSFEG  AM+
Sbjct: 1   MKERVVWGGVAVAIFLPFLIIGNLPFQLFVGVLAMIGVSELLKMKRLEVFSFEGVFAMLA 60

Query: 61  AFVLTVPLDSYLSFLPVDASLSAYGIVIFMILAGTVLNSNSYSFEDAAFPIASSFYVGIG 120
           AFVL VP+D YL+FLP+DA+++ Y +++F ILAGTVLNS +YSF+DAAFPIA+SFYVGIG
Sbjct: 61  AFVLAVPMDHYLTFLPIDANVAFYSLMVFFILAGTVLNSRAYSFDDAAFPIATSFYVGIG 120

Query: 121 FQNLVSARMAGIDKVLLALFIVWATDIGAYMIGRQFGQRKLLPSVSPNKTIEGSLGGIAS 180
           FQ+L++AR++GIDKV LALFIVWATDIGAY+IGRQFG+RKLLP+VSPNKTIEGSLGGIA
Sbjct: 121 FQHLINARLSGIDKVFLALFIVWATDIGAYLIGRQFGRRKLLPTVSPNKTIEGSLGGIAC 180

Query: 181 AIVVAFFFMLFDKTVYAPHSFLVMLVLVAIFSIFGQFGDLVESSIKRHFGVKDSGKLIPG 240
           A++V+F FM+ D++VYAPH FL MLVLVA+FSIF QFGDLVES++KRHFGVKDSGKLIPG
Sbjct: 181 AVLVSFIFMVIDRSVYAPHHFLTMLVLVALFSIFAQFGDLVESALKRHFGVKDSGKLIPG 240

Query: 241 HGGILDRFDSMIFVFPIMHFFGLF 264
           HGGILDRFDSMIFVFPIMH FGLF
Sbjct: 241 HGGILDRFDSMIFVFPIMHLFGLF 264
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 749

A DNA sequence (GBSx0796) was identified in *S.agalactiae* <SEQ ID 2299> which encodes the amino acid sequence <SEQ ID 2300>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.09   Transmembrane    2 -  18 (   1 -  25)
   INTEGRAL    Likelihood = -9.39   Transmembrane  394 - 410 ( 390 - 415)
   INTEGRAL    Likelihood = -8.01   Transmembrane  181 - 197 ( 173 - 198)
   INTEGRAL    Likelihood = -2.97   Transmembrane  343 - 359 ( 342 - 360)

----- Final Results -----
               bacterial membrane --- Certainty=0.5437(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD47948 GB:AF152237 Eep [Enterococcus faecalis]
 Identities = 229/425 (53%), Positives = 298/425 (69%), Gaps = 9/425 (2%)
```

```
Query: 1   MLGILTFIIIFGVIVVVHEFGHFYFAKKSGILVREFAIGMGPKIFSHIDKEGTTYTIRIL 60
           M  I+TFII+FG++V+VHEFGHFYFAK++GILVREFAIGMGPKIF+H  K+GTTYTIR+L
Sbjct: 1   MKTIITFIIVFGILVLVHEFGHFYFAKRAGILVREFAIGMGPKIFAHRGKDGTTYTIRLL 60

Query: 61  PLGGYVRMAGWGDDKTEIKTGTPASLTLNKEGIVTRINLSGKQLDNTSLPINVTAYDLED 120
           P+GGYVRMAG G+D TEI  G P S+ LN  G V +IN S K     S+P+ V  +DLE
Sbjct: 61  PIGGYVRMAGMGEDMTEITPGMPLSVELNAVGNVVKINTSKKVQLPHSIPMEVVDFDLEK 120

Query: 121 KLTITGLV---LSETKTYSVDHDATIIEEDGTEIRIAPLDMQYQNASVWGRLITNFAGPM 177
           +L I G V     E   Y VDHDATIIE DGTE+RIAPLD+Q+Q+A +  R++TNFAGPM
Sbjct: 121 ELFIKGYVNGNEEEETVYKVDHDATIIESDGTEVRIAPLDVQFQSAKLSQRILTNFAGPM 180

Query: 178 NNFILGLVVFIALAFIQGGVQDLSTNQV-RVSENGPAASAGLKNNDRILQIGSHKVSNWE 236
           NNFILG ++F    F+QGGV DL+TNQ+ +V  NGPAA AGLK ND++L I + K+  +E
Sbjct: 181 NNFILGFILFTLAVFLQGGVTDLNTNQIGQVIPNGPAAEAGLKENDKVLSINNQKIKKYE 240

Query: 237 QLTAAVEKSTRHLEKKQKLALKIKSKEVVKTINVKPQKVDKSYI--IGIMPALKTSFKDK 294
             T  V+K+    EK     ++   KE   T+  + QKV+K  I  +G+ P +KT    K
Sbjct: 241 DFTTIVQKNP---EKPLTFVVERNGKEEQLTVTPEKQKVEKQTIGKVGVYPYMKTDLPSK 297

Query: 295 LLGGLKLAWESFFRILNELKKLIAHFSINKLGGPVALYQASSQAAKNGFVTVLNLMGLIS 354
           L+GG++    S  +I   L  L   FS+NKLGGPV +++ S +A+  G  TV+ LM ++S
Sbjct: 298 LMGGIQDTLNSTTQIFKALGSLFTGFSLNKLGGPVMMFKLSEEASNAGVSTVVFLMAMLS 357

Query: 355 INLGIMNLIPIPALDGGKIVMNILEAIRRKPLKQETETYITLAGVAVMLVLMIAVTWNDI 414
           +NLGI+NL+PIPALDGGKIV+NI+E +R KP+  E E  ITL G   ++VLM+ VTWNDI
Sbjct: 358 MNLGIINLLPIPALDGGKIVLNIIEGVRGKPISPEKEGIITLIGFGFVMVLMVLVTWNDI 417

Query: 415 MRAFF 419
            R FF
Sbjct: 418 QRFFF 422
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2301> which encodes the amino acid sequence <SEQ ID 2302>. Analysis of this protein sequence reveals the following:

```
      Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
     INTEGRAL    Likelihood =-11.41   Transmembrane    2 -  18 (   1 -  25)
     INTEGRAL    Likelihood = -9.77   Transmembrane  394 - 410 ( 390 - 415)
     INTEGRAL    Likelihood = -9.61   Transmembrane  180 - 196 ( 173 - 201)
     INTEGRAL    Likelihood = -2.66   Transmembrane  347 - 363 ( 343 - 363)

----- Final Results -----
               bacterial membrane --- Certainty=0.5564(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD47948 GB:AF152237 Eep [Enterococcus faecalis]
 Identities = 230/427 (53%), Positives = 298/427 (68%), Gaps = 13/427 (3%)

Query: 1   MLGIITFIIIFGILVIVHEFGHFYFAKKSGILVREFAIGMGPKIFSHVDQGGTLYTLRML 60
           M  IITFII+FGILV+VHEFGHFYFAK++GILVREFAIGMGPKIF+H  + GT YT+R+L
Sbjct: 1   MKTIITFIIVFGILVLVHEFGHFYFAKRAGILVREFAIGMGPKIFAHRGKDGTTYTIRLL 60

Query: 61  PLGGYVRMAGWGDDKTEIKTGTPASLTLNEQGFVKRINLSQSKLDPTSLPMHVTGYDLED 120
           P+GGYVRMAG G+D TEI  G P S+ LN  G V +IN S+    P S+PM V  +DLE
Sbjct: 61  PIGGYVRMAGMGEDMTEITPGMPLSVELNAVGNVVKINTSKKVQLPHSIPMEVVDFDLEK 120

Query: 121 QLSITGLV---LEETKTYKVAHDATIVEEDGTEIRIAPLDVQYQNASIGGRLITNFAGPM 177
           +L I G V    EE   YKV HDATI+E DGTE+RIAPLDVQ+Q+A +  R++TNFAGPM
Sbjct: 121 ELFIKGYVNGNEEEETVYKVDHDATIIESDGTEVRIAPLDVQFQSAKLSQRILTNFAGPM 180

Query: 178 NNFILGIVVFILLVFLQGGMPDFSSNHV-RVQENGAAAKAGLRDNDQIVAINGYKVTSWN 236
           NNFILG ++F L VFLQGG+ D ++N + +V  NG AA+AGL++ND++++IN  K+  +
Sbjct: 181 NNFILGFILFTLAVFLQGGVTDLNTNQIGQVIPNGPAAEAGLKENDKVLSINNQKIKKYE 240
```

```
Query: 237 DLTEAVDLATRDLGPSQTIKVTYKSHQRLKTVAVKPQKH-AKTYTI---GVKASLKTGFK 292
           D T  V        P + +    + + + + + V P+K   +  TI   GV   +KT
Sbjct: 241 DFTTIV-----QKNPEKPLTFVVERNGKEEQLTVTPEKQKVEKQTIGKVGVYPYMKTDLP 295

Query: 293 DKLLGGLELAWSRAFTILNALKGLITGFSLNKLGGPVAMYDMSNQAAQNGLESVLSLMAM 352
            KL+GG++   +    I  AL  L TGFSLNKLGGPV M+ +S +A+  G+ +V+ LMAM
Sbjct: 296 SKLMGGIQDTLNSTTQIFKALGSLFTGFSLNKLGGPVMMFKLSEEASNAGVSTVVFLMAM 355

Query: 353 LSINLGIFNLIPIPALDGGKILMNIIEAIRRKPIKQETEAYITLAGVAIMVVLMIAVTWN 412
           LS+NLGI NL+PIPALDGGKI++NIIE +R KPI  E E  ITL G   ++VLM+ VTWN
Sbjct: 356 LSMNLGIINLLPIPALDGGKIVLNIIEGVRGKPISPEKEGIITLIGFGFVMVLMVLVTWN 415

Query: 413 DIMRVFF 419
           DI R FF
Sbjct: 416 DIQRFFF 422
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 306/419 (73%), Positives = 359/419 (85%)

Query: 1   MLGILTFIIIFGVIVVVHEFGHFYFAKKSGILVREFAIGMGPKIFSHIDKEGTTYTIRIL 60
           MLGI+TFIIIFG++V+VHEFGHFYFAKKSGILVREFAIGMGPKIFSH+D+ GT YT+R+L
Sbjct: 1   MLGIITFIIIFGILVIVHEFGHFYFAKKSGILVREFAIGMGPKIFSHVDQGGTLYTLRML 60

Query: 61  PLGGYVRMAGWGDDKTEIKTGTPASLTLNKEGIVTRINLSGKQLDNTSLPINVTAYDLED 120
           PLGGYVRMAGWGDDKTEIKTGTPASLTLN++G V RINLS  +LD TSLP++VT YDLED
Sbjct: 61  PLGGYVRMAGWGDDKTEIKTGTPASLTLNEQGFVKRINLSQSKLDPTSLPMHVTGYDLED 120

Query: 121 KLTITGLVLSETKTYSVDHDATIIEEDGTEIRIAPLDMQYQNASVWGRLITNFAGPMNNF 180
           +L+ITGLVL ETKTY V HDATI+EEDGTEIRIAPLD+QYQNAS+ GRLITNFAGPMNNF
Sbjct: 121 QLSITGLVLEETKTYKVAHDATIVEEDGTEIRIAPLDVQYQNASIGGRLITNFAGPMNNF 180

Query: 181 ILGLVVFIALAFIQGGVQDLSTNQVRVSENGPAASAGLKNNDRILQIGSHKVSNWEQLTA 240
           ILG+VVFI L F+QGG+ D S+N VRV ENG AA AGL++ND+I+ I  +KV++W  LT
Sbjct: 181 ILGIVVFILLVFLQGGMPDFSSNHVRVQENGAAAKAGLRDNDQIVAINGYKVTSWNDLTE 240

Query: 241 AVEKSTRHLEKKQKLALKIKSKEVVKTINVKPQKVDKSYIIGIMPALKTSFKDKLLGGLK 300
           AV+ +TR L   Q + +  KS + +KT+ VKPQK  K+Y IG+  +LKT FKDKLLGGL+
Sbjct: 241 AVDLATRDLGPSQTIKVTYKSHQRLKTVAVKPQKHAKTYTIGVKASLKTGFKDKLLGGLE 300

Query: 301 LAWESFFRILNELKKLIAHFSINKLGGPVALYQASSQAAKNGFVTVLNLMGLISINLGIM 360
           LAW   F ILN LK LI  FS+NKLGGPVA+Y  S+QAA+NG  +VL+LM ++SINLGI
Sbjct: 301 LAWSRAFTILNALKGLITGFSLNKLGGPVAMYDMSNQAAQNGLESVLSLMAMLSINLGIF 360

Query: 361 NLIPIPALDGGKIVMNILEAIRRKPLKQETETYITLAGVAVMLVLMIAVTWNDIMRAFF 419
           NLIPIPALDGGKI+MNI+EAIRRKP+KQETE YITLAGVA+M+VLMIAVTWNDIMR FF
Sbjct: 361 NLIPIPALDGGKILMNIIEAIRRKPIKQETEAYITLAGVAIMVVLMIAVTWNDIMRVFF 419
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 750

A DNA sequence (GBSx0797) was identified in *S.agalactiae* <SEQ ID 2303> which encodes the amino acid sequence <SEQ ID 2304>. This protein is predicted to be prolyl-tRNA synthetase (proS). Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32    Transmembrane  473 - 489 ( 473 - 490)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10181> which encodes amino acid sequence <SEQ ID 10182> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13530 GB:Z99112 prolyl-tRNA synthetase [Bacillus subtilis]
 Identities = 301/608 (49%), Positives = 410/608 (66%), Gaps = 52/608 (8%)

Query: 1   MKQSKMLIPTLREMPSDAQVISHALMVRAGYVRQVSAGIYAYLPLANRTIEKFKTIMRQE 60
            M+QS  LIPTLRE+P+DA+  SH L++RAG++RQ ++G+Y+Y+PLA + I+  + I+R+E
Sbjct: 1   MRQSLTLIPTLREVPADAEAKSHQLLLRAGFIRQNTSGVYSYMPLAYKVIQNIQQIVREE 60

Query: 61  FEKIGAVEMLAPALLTADLWRESGRYETYGEDLYKLKNRDQSDFILGPTHEETFTTLVRD 120
             EKI AVEML PAL  A+ W+ESGR+ TYG +L +LK+R   +F LG THEE  T+LVRD
Sbjct: 61  MEKIDAVEMLMPALQQAETWQESGRWYTYGPELMRLKDRHGREFALGATHEEVITSLVRD 120

Query: 121 AVKSYKQLPLNLYQIQSKYRDEKRPRNGLLRTREFIMKDGYSFHKDYEDLDVTYEDYRKA 180
             VKSYK+LPL LYQIQSK+RDEKRPR GLLR REFIMKD YSFH   E LD TY+   +A
Sbjct: 121 EVKSYKRLPLTLYQIQSKFRDEKRPRFGLLRGREFIMKDAYSFHASAESLDETYQKMYEA 180

Query: 181 YEAIFTRAGLDFKGIIGDGGAMGGKDSQEFMAVTPNRTDLNRWLVLDKTIPSIDDIPEDV 240
            Y  IF R G++ + +I D GAMGGKD+ EFMA++
Sbjct: 181 YSNIFARCGINVRPVIADSGAMGGKDTHEFMALS-------------------------- 214

Query: 241 LEEIKVELSAWLVSGEDTIAYSTESSYAANLEMATNEYKPSTKAATFEEVTKVETPNCKS 300
                          GEDTIAYS ES YAAN+EMA   ++        + + KV TPN K+
Sbjct: 215 ------------AIGEDTIAYSDESQYAANIEMAEVLHQEVPSDEEPKALEKVHTPNVKT 262

Query: 301 IDEVAGFLSIDENQTIKTLLFIADEQPVVALLVGNDQVNDVKLKNYLAADFLEPASEEQA 360
            I+E+  FL +     IK++LF AD++ V+ L+ G+ +VND+K+KN L A+ +E A+ E+
Sbjct: 263 IEELTAFLQVSAEACIKSVLFKADDRFVLVLVRGDHEVNDIKVKNLLHAEVVELATHEEV 322

Query: 361 KEIFGAGFGSLGPVNLPDSVKIIADRKVQDLANAVSGANQDGYHFTGVNPERDFTA-EYV 419
             +  G   G +GPV +   V++ AD+ V+ + NAV+GAN+  +H+  VN  RD    E+
Sbjct: 323 IQQLGTEPGFVGPVGIHQDVEVYADQAVKAMVNAVAGANEGDHHYKNVNVNRDAQIKEFA 382

Query: 420 DIREVKEGEISPDGKGTLKFARGIEIGHIFKLGTRYSDSMGANILDENGRSNPIVMGCYG 479
            D+R +KEG+ SPDGKGT++FA GIE+G +FKLGTRYS++M A  LDENGR+ P++MGCYG
Sbjct: 383 DLRFIKEGDPSPDGKGTIRFAEGIEVGQVFKLGTRYSEAMNATYLDENGRAQPMLMGCYG 442

Query: 480 IGVSRILSAVIEQHARLFVNKTPKGAYRFAWGINFPEELAPFDVHLITVNVKDQESQDLT 539
            IGVSR LSA+ EQH                 G+ +P+ +AP+D+H++ +N+K+   ++L
Sbjct: 443 IGVSRTLSAIAEQH-------------HDEKGLIWPKSVAPYDLHILALNMKNDGQRELA 489

Query: 540 EKIEADLMLKGYEVLTDDRNERVGSKFSDSDLIGLPIRVTVGKKASEGIVEVKIKASGDT 599
            EK+ ADL  +GYEVL DDR ER G KF+DSDLIGLPIR+TVGK+A EGIVEVKI+ +G++
Sbjct: 490 EKLYADLKAEGYEVLYDDRAERAGVKFADSDLIGLPIRITVGKRADEGIVEVKIRQTGES 549

Query: 600 IEVHADNL 607
             E+  D L
Sbjct: 550 TEISVDEL 557
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2305> which encodes the amino acid sequence <SEQ ID 2306>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane  473 - 489 ( 473 - 490)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 535/617 (86%), Positives = 584/617 (93%)

Query: 1   MKQSKMLIPTLREMPSDAQVISHALMVRAGYVRQVSAGIYAYLPLANRTIEKFKTIMRQE 60
            MKQSK+LIPTLREMPSDAQVISHALMVRAGYVRQVSAGIYAYLPLANRTIEKFKTIMR+E
Sbjct: 1   MKQSKLLIPTLREMPSDAQVISHALMVRAGYVRQVSAGIYAYLPLANRTIEKFKTIMREE 60

Query: 61  FEKIGAVEMLAPALLTADLWRESGRYETYGEDLYKLKNRDQSDFILGPTHEETFTTLVRD 120
            FEKIGAVEMLAPALLTADLWRESGRYETYGEDLYKLKNRD SDFILGPTHEETFTTLVRD
Sbjct: 61  FEKIGAVEMLAPALLTADLWRESGRYETYGEDLYKLKNRDNSDFILGPTHEETFTTLVRD 120

Query: 121 AVKSYKQLPLNLYQIQSKYRDEKRPRNGLLRTREFIMKDGYSFHKDYEDLDVTYEDYRKA 180
            AVKSYKQLPLNLYQIQSKYRDEKRPRNGLLRTREFIMKDGYSFH +YEDLDVTYEDYR+A
Sbjct: 121 AVKSYKQLPLNLYQIQSKYRDEKRPRNGLLRTREFIMKDGYSFHHNYEDLDVTYEDYRQA 180

Query: 181 YEAIFTRAGLDFKGIIGDGGAMGGKDSQEFMAVTPNRTDLNRWLVLDKTIPSIDDIPEDV 240
            YEAIFTRAGLDFKGIIGDGGAMGGKDSQEFMA+TP RTDL+RW+VLDK+I S+DDIP++V
Sbjct: 181 YEAIFTRAGLDFKGIIGDGGAMGGKDSQEFMAITPARTDLDRWVVLDKSIASMDDIPKEV 240

Query: 241 LEEIKVELSAWLVSGEDTIAYSTESSYAANLEMATNEYKPSTKAATFEEVTKVETPNCKS 300
            LE+IK EL+AW++SGEDTIAYSTESSYAANLEMATNEYKPS+K A  + + +VETP+CK+
Sbjct: 241 LEDIKAELAAWMISGEDTIAYSTESSYAANLEMATNEYKPSSKVAAEDALAEVETPHCKT 300

Query: 301 IDEVAGFLSIDENQTIKTLLFIADEQPVVALLVGNDQVNDVKLKNYLAADFLEPASEEQA 360
            IDEVA FLS+DE QTIKTLLF+AD +PVVALLVGND +N VKLKNYLAADFLEPASEE+A
Sbjct: 301 IDEVAAFLSVDETQTIKTLLFVADNEPVVALLVGNDHINTVKLKNYLAADFLEPASEEEA 360

Query: 361 KEIFGAGFGSLGPVNLPDSVKIIADRKVQDLANAVSGANQDGYHFTGVNPERDFTAEYVD 420
            +  FGAGFGSLGPVNL    +I+ADRKVQ+L NAV+GAN+DG+H TGVNP RDF AEYVD
Sbjct: 361 RAFFGAGFGSLGPVNLAQGSRIVADRKVQNLTNAVAGANKDGFHMTGVNPGRDFQAEYVD 420

Query: 421 IREVKEGEISPDGKGTLKFARGIEIGHIFKLGTRYSDSMGANILDENGRSNPIVMGCYGI 480
            IREVKEGE+SPDG G L+FARGIE+GHIFKLGTRYSDSMGA ILDENGR+ PIVMGCYGI
Sbjct: 421 IREVKEGEMSPDGHGVLQFARGIEVGHIFKLGTRYSDSMGATILDENGRTVPIVMGCYGI 480

Query: 481 GVSRILSAVIEQHARLFVNKTPKGAYRFAWGINFPEELAPFDVHLITVNVKDQESQDLTE 540
            GVSRILSAVIEQHARLFVNKTPKG YR+AWGINFP+ELAPFDVHLITVNVKDQ +QDLT
Sbjct: 481 GVSRILSAVIEQHARLFVNKTPKGDYRYAWGINFPKELAPFDVHLITVNVKDQVAQDLTA 540

Query: 541 KIEADLMLKGYEVLTDDRNERVGSKFSDSDLIGLPIRVTVGKKASEGIVEVKIKASGDTI 600
            K+EADLM KGY+VLTDDRNERVGSKFSDSDLIGLPIRVTVGKKA+EGIVE+KIKA+GD+I
Sbjct: 541 KLEADLMAKGYDVLTDDRNERVGSKFSDSDLIGLPIRVTVGKKAAEGIVEIKIKATGDSI 600

Query: 601 EVHADNLIETLEILTKK 617
            EV+A+NLIETLEILTK+
Sbjct: 601 EVNAENLIETLEILTKE 617
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 751

A DNA sequence (GBSx0798) was identified in *S.agalactiae* <SEQ ID 2307> which encodes the amino acid sequence <SEQ ID 2308>. This protein is predicted to be peptidoglycan hydrolase (flgJ). Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.86    Transmembrane    9 -  25 (   9 -  25)

----- Final Results -----
               bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB94815 GB:AJ245582 peptidoglycan hydrolase [Streptococcus thermophilus]
 Identities = 101/201 (50%), Positives = 122/201 (60%), Gaps = 9/201 (4%)

Query: 2   KSRKKDKLVLRLTT-----TLLVFGL----GGVWFYNYKNDNVEPTVTSASDQTTTFIQT 52
            KS+KK K VL         +L+  GL    G +   N+    +E  +T   + T  FI
Sbjct: 16  KSKKKKKSVLLFPKFFQKWSLIFIGLFSLLGLLASLNFPRLTMEKNMTPTDETTVAFIAE 75

Query: 53  ISPTAIEISKTYDLYASVLLAQAILESSSGQSDLSKAPNYNLFGIKGEYKGKSVQMPTLE 112
            I  T+  ++   DLYASV++AQAILES SGQS LS+ P YN FGIKGEY G+SV +PT E
Sbjct: 76  IGETSRYLAARNDLYASVMIAQAILESDSGQSQLSQKPLYNFFGIKGEYNGQSVTLPTWE 135

Query: 113 DDGKGNMTQIQAPFRAYPNYSASLYDYAELVSSQKYASVWKSNTSSYKDATAALTGLYAT 172
            DDGKGN   I A FR+Y +   SL DY E +    Y  V +S T SYKDATAALTG+YAT
Sbjct: 136 DDGKGNPYHIDAAFRSYGSVENSLQDYVEFLEGSYYVGVHRSKTRSYKDATAALTGVYAT 195

Query: 173 DTAYASKLNQIIETYSLDAYD 193
            DT Y  KLN IIE Y L  YD
Sbjct: 196 DTTYGDKLNSIIEQYQLTIYD 216
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2309> which encodes the amino acid sequence <SEQ ID 2310>. Analysis of this protein sequence reveals the following:

```
     Possible site: 24

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB94815 GB:AJ245582 peptidoglycan hydrolase [Streptococcus thermophilus]
 Identities = 103/189 (54%), Positives = 126/189 (66%), Gaps = 4/189 (2%)

Query: 4   KKGKLVLISLFVLAACLGAYSAMRQSHKTSNVSAETIASSSTRHFIDEIGPTASTIGQER 63
            +K  L+ I LF L   L + +  R + + +     T    +T  FI EIG T+   +
Sbjct: 32  QKWSLIFIGLFSLLGLLASLNFPRLTMEKNM----TPTDETTVAFIAEIGETSRYLAARN 87

Query: 64  DLYASVMIAQAILESSNGKSSLSQAPYYNFFGIKGAYNGSSVTMSTWEDDGNGNTYTIDQ 123
            DLYASVMIAQAILES +G+S LSQ P YNFFGIKG YNG SVT+ TWEDDG GN Y ID
Sbjct: 88  DLYASVMIAQAILESDSGQSQLSQKPLYNFFGIKGEYNGQSVTLPTWEDDGKGNPYHIDA 147

Query: 124 AFRAYPSIADSLNDYADLLSSSTYIGARKSNTLSYQDATAALTGLYATDTSYNLKLNNII 183
            AFR+Y S+ +SL DY + L  S Y+G  +S T SY+DATAALTG+YATDT+Y  KLN+II
Sbjct: 148 AFRSYGSVENSLQDYVEFLEGSYYVGVHRSKTRSYKDATAALTGVYATDTTYGDKLNSII 207

Query: 184 ATYGLTAYD 192
              Y LT YD
Sbjct: 208 EQYQLTIYD 216
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 108/192 (56%), Positives = 124/192 (64%), Gaps = 2/192 (1%)

Query: 3   SRKKDKLVL-RLTTTLLVFGLGGVWFYNYKNDNVEPTVTSASDQTTTFIQTISPTAIEIS 61
            ++KK KLVL  L       G       ++K  NV    T AS  T  FI  I PTA  I
Sbjct: 2   TKKKGKLVLISLFVLAACLGAYSAMRQSHKTSNVSAE-TIASSSTRHFIDEIGPTASTIG 60

Query: 62  KTYDLYASVLLAQAILESSSGQSDLSKAPNYNLFGIKGEYKGKSVQMPTLEDDGKGNMTQ 121
            +  DLYASV++AQAILESS+G+S LS+AP YN FGIKG Y G SV M T EDDG GN
Sbjct: 61  QERDLYASVMIAQAILESSNGKSSLSQAPYYNFFGIKGAYNGSSVTMSTWEDDGNGNTYT 120

Query: 122 IQAPFRAYPNYSASLYDYAELVSSQKYASVWKSNTSSYKDATAALTGLYATDTAYASKLN 181
```

```
           I   FRAYP+ + SL DYA+L+SS  Y    KSNT SY+DATAALTGLYATDT+Y  KLN
Sbjct: 121 IDQAFRAYPSIADSLNDYADLLSSSTYIGARKSNTLSYQDATAALTGLYATDTSYNLKLN 180

Query: 182 QIIETYSLDAYD 193
            II TY L AYD
Sbjct: 181 NIIATYGLTAYD 192
```

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9073> which encodes the amino acid sequence <SEQ ID 9074>. Analysis of this protein sequence reveals the following:

```
     Possible site: 58

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS sequences follows:

```
 Score =  130 bits (323), Expect = 2e-32
 Identities = 68/169 (40%), Positives = 96/169 (56%), Gaps = 3/169 (1%)

Query: 30  MWTLKLGNQRLAPY---ADHETLTFVRKISHAAQSVAQKKQLYSSVMMAQAILESNNGKS 86
            +W     N  + P    A  +T TF++ IS  A  +++   LY+SV++AQAILES++G+S
Sbjct: 25  VWFYNYKNDNVEPTVTSASDQTTTFIQTISPTAIEISKTYDLYASVLLAQAILESSSGQS 84

Query: 87  QLSQKPYYNFFGIKGSYKERSVIFPTLEDDGQGNLYQIDAAFRSYGSLTACFLDYARVLN 146
             LS+ P YN FGIKG YK +SV  PTLEDDG+GN+ QI A FR+Y + +A   DYA +++
Sbjct: 85  DLSKAPNYNLFGIKGEYKGKSVQMPTLEDDGKGNMTQIQAPFRAYPNYSASLYDYAELVS 144

Query: 147 DPLYDKTHKKFWSHYQXXXXXXXXXXXXXXXXXXKLNELIEWYQLTNFD 195
               Y    K   S Y+                  KLN++IE Y L  +D
Sbjct: 145 SQKYASVWKSNTSSYKDATAALTGLYATDTAYASKLNQIIETYSLDAYD 193
```

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9075> which encodes the amino acid sequence <SEQ ID 9076>. An alignment of the GAS and GBS sequences follows:

```
 Score = 69.1 bits (166), Expect = 1e-13
 Identities = 52/151 (34%), Positives = 79/151 (51%), Gaps = 10/151 (6%)

Query: 2   TFLDKIKQGCLDGWAKYKILPSLTAAQAILESGWGKH----APHNALFGIKADSSWTGKS 57
            TF+  I    ++    Y +  S+  AQAILES  G+     AP+  LFGIK +  + GKS
Sbjct: 48  TFIQTISPTAIEISKTYDLYASVLLAQAILESSSGQSDLSKAPNYNLFGIKGE--YKGKS 105

Query: 58  FDTKTQEEYQAGVVTDIVDRFRAYDSWDESIADHGQFLVDNPRYEAV--IGETDYKKACY 115
                T E+   G +T I   FRAY ++  S+ D+ + LV + +Y +V     + YK A
Sbjct: 106 VQMPTLEDDGKGNMTQIQAPFRAYPNYSASLYDYAE-LVSSQKYASVWKSNTSSYKDATA 164

Query: 116 AIKAAGYATASSYVELLIQLIEENDLQSWDR 146
            A+    YAT ++Y   L Q+IE   L ++D+
Sbjct: 165 ALTGL-YATDTAYASKLNQIIETYSLDAYDK 194
```

SEQ ID 2308 (GBS275) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 4; MW 22.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 4; MW 47.5kDa).

The GBS275-GST fusion product was purified (Figure 208, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 276), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 752

A DNA sequence (GBSx0799) was identified in *S.agalactiae* <SEQ ID 2311> which encodes the amino acid sequence <SEQ ID 2312>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16    Transmembrane  876 - 892 ( 876 - 892)

----- Final Results -----
               bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2313> which encodes the amino acid sequence <SEQ ID 2314>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16    Transmembrane  873 - 889 ( 873 - 889)

----- Final Results -----
               bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB94815 GB:AJ245582 peptidoglycan hydrolase [Streptococcus thermophilus]
 Identities = 96/202 (47%), Positives = 127/202 (62%), Gaps = 10/202 (4%)

Query: 4   KKRRRRAKSSV---------NRLVLGLV-LLNLIVSMWTLKLGNQRLAPYADHETLTFVR 53
           KK +++ KS +         + + +GL  LL L+ S+   +L  ++     D  T+ F+
Sbjct: 15  KKSKKKKKSVLLFPKFFQKWSLIFIGLFSLLGLLASLNFPRLTMEKNMTPTDETTVAFIA 74

Query: 54  KISHAAQSVAQKKQLYSSVMMAQAILESNNGKSQLSQKPYYNFFGIKGSYKERSVIFPTL 113
           +I   ++ +A +  LY+SVM+AQAILES++G+SQLSQKP YNFFGIKG Y  +SV  PT
Sbjct: 75  EIGETSRYLAARNDLYASVMIAQAILESDSGQSQLSQKPLYNFFGIKGEYNGQSVTLPTW 134

Query: 114 EDDGQGNLYQIDAAFRSYGSLTACFLDYARVLNDPLYDKTHKKFWSHYQDATATLTGTYA 173
           EDDG+GN Y IDAAFRSYGS+     DY   L    Y   H+     Y+DATA LTG YA
Sbjct: 135 EDDGKGNPYHIDAAFRSYGSVENSLQDYVEFLEGSYYVGVHRSKTRSYKDATAALTGVYA 194

Query: 174 TDTTYHTKLNELIEWYQLTNFD 195
           TDTTY  KLN +IE YQLT +D
Sbjct: 195 TDTTYGDKLNSIIEQYQLTIYD 216
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 1244/1468 (84%), Positives = 1351/1468 (91%), Gaps = 3/1468 (0%)

Query: 1    MSELFKKLMDQIEMPLEIKNSSVFSSADIIEVKVHSLSRLWEFHFSFPELLPIEVYRELQ 60
            MS+LF KLMDQIEMPL+++ SS FSSADIIEVKVHS+SRLWEFHF+F  +LPI  YREL
Sbjct: 1    MSDLFAKLMDQIEMPLDMRRSSAFSSADIIEVKVHSVSRLWEFHFAFAAVLPIATYRELH 60

Query: 61   TRLVNSFEKADIKATFDIRAETIDFSDDLLQDYYQQAFCEPLCNSASFKSSFSQLKVHYN 120
             RL+ +FE ADIK TFDI+A  +D+SDDLLQ YYQ+AF    CNSASFKSSFS+LKV Y
Sbjct: 61   DRLIRTFEAADIKVTFDIQAAQVDYSDDLLQAYYQEAFEHAPCNSASFKSSFSKLKVTYE 120

Query: 121  GSQMIISAPQFVNNNHFRQNHLPRLEQQFSLFGFGKLAIDMVSDEQMTQDLKSSFETNRE 180
              ++II+AP FVNN+HFR NHLP L +Q   FGFG L IDMVSD++MT+ L  +F ++R+
Sbjct: 121  DDKLIIAAPGFVNNDHFRNNHLPNLVKQLEAFGFGILTIDMVSDQEMTEHLTKNFVSSRQ 180
```

```
Query: 181  QLLEKANQEAMQALEAQKSLEDSAPPSEEVTPTQNYDFKERIKQRQAGFEKAEITPMIEV 240
             L++KA Q+    LEAQKSLE   PP EE TP   +D+KER  +RQAGFEKA ITPMIE+
Sbjct: 181  ALVKKAVQDN---LEAQKSLEAMMPPVEEATPAPKFDYKERAAKRQAGFEKATITPMIEI 237

Query: 241  TTEENRIVFEGMVFSVERKTTRTGRHIINFKMTDYTSSFAMQKWAKDDEELKKYDMISKG 300
             TEENRIVFEGMVF VERKTTRTGRHIINFKMTDYTSSFA+QKWAKDDEEL+K+DMI+KG
Sbjct: 238  ETEENRIVFEGMVFDVERKTTRTGRHIINFKMTDYTSSFALQKWAKDDEELRKFDMIAKG 297

Query: 301  SWLRVRGNIENNNFTKSLTMNVQDIKEIVHHERKDLMPADQKRVEFHAHTNMSTMDALPT 360
            +WLRV+GNIE N FTKSLTMNVQ +KEIV HERKDLMP  QKRVE HAHTNMSTMDALPT
Sbjct: 298  AWLRVQGNIETNPFTKSLTMNVQQVKEIVRHERKDLMPEGQKRVELHAHTNMSTMDALPT 357

Query: 361  VESLIDTAAKWGHPAIAITDHANVQSFPHGYHRAKKAGIKAIFGLEANIVEDKVPISYNE 420
            VESLIDTAAKWGH AIAITDHANVQSFPHGYHRA+KAGIKAIFGLEANIVEDKVPISY
Sbjct: 358  VESLIDTAAKWGHKAIAITDHANVQSFPHGYHRARKAGIKAIFGLEANIVEDKVPISYEP 417

Query: 421  VDMNLHEATYVVFDVETTGLSAANNDLIQIAASKMFKGNIIEQFDEFIDPGHPLSAFTTE 480
            VDM+LHEATYVVFDVETTGLSA NNDLIQIAASKMFKGNI+EQFDEFIDPGHPLSAFTTE
Sbjct: 418  VDMDLHEATYVVFDVETTGLSAMNNDLIQIAASKMFKGNIVEQFDEFIDPGHPLSAFTTE 477

Query: 481  LTGITDNHVRGSKPILQVLQEFQNFCQGTVLVAHNATFDVGFMNANYERHNLPLITQPVI 540
            LTGITD H++G+KP++ VL+ FQ+FC+ ++LVAHNA+FDVGFMNANYERH+LP ITQPVI
Sbjct: 478  LTGITDKHLQGAKPLVTVLKAFQDFCKDSILVAHNASFDVGFMNANYERHDLPKITQPVI 537

Query: 541  DTLEFARNLYPEYKRHGLGPLTKRFQVALEHHHMANYDAEATGRLLFIFLKEARENRDVT 600
            DTLEFARNLYPEYKRHGLGPLTKRFQV+L+HHHMANYDAEATGRLLFIFLK+ARE   +
Sbjct: 538  DTLEFARNLYPEYKRHGLGPLTKRFQVSLDHHHMANYDAEATGRLLFIFLKDAREKHGIK 597

Query: 601  NLMELNTKLVAEDSYKKARIKHATIYVQNQVGLKNIFKLVSLSNVKYFEGVARIPRSVLD 660
            NL++LNT LVAEDSYKKARIKHATIYVQNQVGLKN+FKLVSLSN+KYFEGV RIPR+VLD
Sbjct: 598  NLLQLNTDLVAEDSYKKARIKHATIYVQNQVGLKNMFKLVSLSNIKYFEGVPRIPRTVLD 657

Query: 661  AHREGLLLGTACSDGEVFDALLSNGIDAAVTLAKYYDFIEVMPPAIYRPLVVRDLIKDEV 720
            AHREGLLLGTACSDGEVFDA+L+ GIDAAV LA+YYDFIE+MPPAIY+PLVVR+LIKD+
Sbjct: 658  AHREGLLLGTACSDGEVFDAVLTKGIDAAVDLARYYDFIEIMPPAIYQPLVVRELIKDQA 717

Query: 721  GIQQIIRDLIEVGRRLDKPVLATGNVHYIEPEDEIYREIIVRSLGQGAMINRTIGRGEDA 780
            GI+Q+IRDLIEVG+R  KPVLATGNVHY+EPE+EIYREIIVRSLGQGAMINRTIGRGE A
Sbjct: 718  GIEQVIRDLIEVGKRAKKPVLATGNVHYLEPEEEIYREIIVRSLGQGAMINRTIGRGEGA 777

Query: 781  QPAPLPKAHFRTTNEMLDEFAFLGKDLAYEIVVTNTNTFADRFEDVEVVKGDLYTPFVDR 840
            QPAPLPKAHFRTTNEMLDEFAFLGKDLAY++VV NT  FADR E+VEVVKGDLYTP++D+
Sbjct: 778  QPAPLPKAHFRTTNEMLDEFAFLGKDLAYQVVVQNTQDFADRIEEVEVVKGDLYTPYIDK 837

Query: 841  AEERVAELTYAKAFEIYGNPLPDIIDLRIEKELASILGNGFAVIYLASQMLVQRSNERGY 900
            AEE VAELTY KAFEIYGNPLPDIIDLRIEKEL SILGNGFAVIYLASQMLV RSNERGY
Sbjct: 838  AEETVAELTYQKAFEIYGNPLPDIIDLRIEKELTSILGNGFAVIYLASQMLVNRSNERGY 897

Query: 901  LVGSRGSVGSSFVATMIGITEVNPMPPHYVCPNCQHSEFITDGSCGSGYDLPNKNCPKCG 960
            LVGSRGSVGSSFVATMIGITEVNPMPPHYVCP+CQHSEFITDGS GSGYDLPNK CPKCG
Sbjct: 898  LVGSRGSVGSSFVATMIGITEVNPMPPHYVCPSCQHSEFITDGSVGSGYDLPNKPCPKCG 957

Query: 961  TLYKKDGQDIPFETFLGFDGDKVPDIDLNFSGDDQPSAHLDVRDIFGEEYAFRAGTVGTV 1020
            T Y+KDGQDIPFETFLGFDGDKVPDIDLNFSGDDQPSAHLDVRDIFG+EYAFRAGTVGTV
Sbjct: 958  TPYQKDGQDIPFETFLGFDGDKVPDIDLNFSGDDQPSAHLDVRDIFGDEYAFRAGTVGTV 1017

Query: 1021 AEKTAFGFVKGYERDYNKFYNDAEVERLATGAAGVKRSTGQHPGGIVVIPNYMDVYDFTP 1080
            AEKTA+GFVKGYERDY KFY DAEV+RLA GAAGVKR+TGQHPGGIVVIPNYMDVYDFTP
Sbjct: 1018 AEKTAYGFVKGYERDYGKFYRDAEVDRLAAGAAGVKRTTGQHPGGIVVIPNYMDVYDFTP 1077

Query: 1081 VQYPADDMTAAWQTTHFNFHDIDENVLKLDILGHDDPTMIRKLQDLSGIDPSNILPDDPD 1140
            VQYPADD+TA+WQTTHFNFHDIDENVLKLDILGHDDPTMIRKLQDLSGIDP  I  DDP
Sbjct: 1078 VQYPADDVTASWQTTHFNFHDIDENVLKLDILGHDDPTMIRKLQDLSGIDPITIPADDPG 1137

Query: 1141 VMKLFSGTEVLGVTEEQIGTPTGMLGIPEFGTNFVRGMVNETHPTTFAELLQLSGLSHGT 1200
            VM LFSGTEVLGVT EQIGTPTGMLGIPEFGTNFVRGMVNETHPTTFAELLQLSGLSHGT
Sbjct: 1138 VMALFSGTEVLGVTPEQIGTPTGMLGIPEFGTNFVRGMVNETHPTTFAELLQLSGLSHGT 1197
```

```
Query: 1201 DVWLGNAQDLIKEGIATLSTVIGCRDDIMVYLMHAGLQPKMAFTIMERVRKGLWLKISED 1260
            DVWLGNAQDLIKEGIATL TVIGCRDDIMVYLMHAGL+PKMAFTIMERVRKGLWLKISE+
Sbjct: 1198 DVWLGNAQDLIKEGIATLKTVIGCRDDIMVYLMHAGLEPKMAFTIMERVRKGLWLKISEE 1257

Query: 1261 ERNGYIQAMRDNNVPDWYIESCGKIKYMFPKAHAAAYVLMALRVAYFKVHYPIFYYCAYF 1320
            ERNGYI AMR+NNVPDWYIESCGKIKYMFPKAHAAAYVLMALRVAYFKVH+PI YYCAYF
Sbjct: 1258 ERNGYIDAMRENNVPDWYIESCGKIKYMFPKAHAAAYVLMALRVAYFKVHHPIMYYCAYF 1317

Query: 1321 SIRAKAFELRTMSAGLDAVKARMKDITEKRQRNEATNVENDLFTTLELVNEMLERGFKFG 1380
            SIRAKAFEL+TMS GLDAVKARM+DIT KR+ NEATNVENDLFTTLE+VNEMLERGFKFG
Sbjct: 1318 SIRAKAFELKTMSGGLDAVKARMEDITIKRKNNEATNVENDLFTTLEIVNEMLERGFKFG 1377

Query: 1381 KLDLYRSHATDFIIEEDTLIPPFVAMEGLGENVAKQIVRAREDGEFLSKTELRKRGGVSS 1440
            KLDLY+S A +F I+ DTLIPPF+A+EGLGENVAKQIV+AR++GEFLSK ELRKRGG SS
Sbjct: 1378 KLDLYKSDAIEFQIKGDTLIPPFIALEGLGENVAKQIVKARQEGEFLSKMELRKRGGASS 1437

Query: 1441 TLVEKFDEMGILGNLPEDNQLSLFDDFF 1468
            TLVEK DEMGILGN+PEDNQLSLFDDFF
Sbjct: 1438 TLVEKMDEMGILGNMPEDNQLSLFDDFF 1465
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 753

A DNA sequence (GBSx0800) was identified in *S.agalactiae* <SEQ ID 2315> which encodes the amino acid sequence <SEQ ID 2316>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1505(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10179> which encodes amino acid sequence <SEQ ID 10180> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13207 GB:Z99111 similar to transcriptional regulator (MarR
           family) [Bacillus subtilis]
 Identities = 49/124 (39%), Positives = 73/124 (58%)

Query: 18  VMRKAFRTIDGKVSESFKEFELTPTQFAVLDVLYAKGTMKIGELIENMLATSGNMTVVIK 77
           V  +AF+++        KE    PT+FAVL++LY +G  K+ ++   +L  SGN+T VI
Sbjct: 20  VFARAFKSVSEHSIRDSKEHGFNPTEFAVLELLYTRGPQKLQQIGSRLLLVSGNVTYVID 79

Query: 78  NMEKKGWVLRHSCPNDKRAFLVSLTTEGEEVIKKALPEHIKRVEDAFSVLTETEQEDLIN 137
            +E+ G+++R   P DKR+    LT +G E + K  P H  R+  AFS L+  EQ+ LI
Sbjct: 80  KLERNGFLVREQDPKDKRSVYAHLTDKGNEYLDKIYPIHALRIARAFSGLSPDEQDQLIV 139

Query: 138 LLKK 141
           LLKK
Sbjct: 140 LLKK 143
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2317> which encodes the amino acid sequence <SEQ ID 2318>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0537(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 80/145 (55%), Positives = 111/145 (76%), Gaps = 1/145 (0%)

Query: 2   GDEMGNF-KNSAVKSMVVMRKAFRTIDGKVSESFKEFELTPTQFAVLDVLYAKGTMKIGE 60
           G++M +  KN+A+K+MVV RKA RT+D   ++ FK+ +LT TQF+VL+VLY KG M+I
Sbjct: 8   GNQMSHLDKNTALKAMVVFRKAQRTLDAFGADIFKKADLTATQFSVLEVLYTKGCMRINH 67

Query: 61  LIENMLATSGNMTVVIKNMEKKGWVLRHSCPNDKRAFLVSLTTEGEEVIKKALPEHIKRV 120
           LI+++LATSGNMTVV+ NME+ GW+ +     DKRA++V+LT +G  +I+  LP+H+ RV
Sbjct: 68  LIDSLLATSGNMTVVLNNMERNGWISKCKDKTDKRAYVVTLTDKGTRLIEAVLPKHVARV 127

Query: 121 EDAFSVLTETEQEDLINLLKKFKTL 145
           E+AF+VLTE EQ  LI LLKKFK L
Sbjct: 128 EEAFAVLTEKEQLCLIELLKKFKQL 152
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 754

A DNA sequence (GBSx0801) was identified in *S.agalactiae* <SEQ ID 2319> which encodes the amino acid sequence <SEQ ID 2320>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3742(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG05963 GB:AE004686 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 115/203 (56%), Positives = 143/203 (69%), Gaps = 7/203 (3%)

Query: 2   SFLEELKNRRSIYALGRNTEVSDEKIVEIIKEAVRQSPSAFNSQTSRVVILLNDEVTKFW 61
           +FL  +KNRR+IYAL +   VS EKIVE++KEAV  SPSAFNSQ+SRVV+L   E  +FW
Sbjct: 4   AFLSSIKNRRTIYALDKQLPVSQEKIVELVKEAVSHSPSAFNSQSSRVVVLFGAEHEQFW 63

Query: 62  DELVANDLVETMKVQGAPETAIAGTKEKLASFGASKGTVLFFEDQDVVKSLQEQFVLYAD 121
           +  +A D  E  K+   P  A A T+ KL SF A  GTVLFFEDQ VV+ LQEQF LYAD
Sbjct: 64  N--IAKD--ELKKI--VPADAFAATETKLNSFAAGAGTVLFFEDQTVVRQLQEQFALYAD 117

Query: 122 NFPVWSEQSTGIASVNTWTALSAELGLGGNLQHYNPVIDASVQAVYGVPASWKLRGQLNF 181
           NFPVWSEQ++G+A    WTAL AE  +G +LQHYNP++DA     + +P SWKLR Q+ F
Sbjct: 118 NFPVWSEQASGMAQFAVWTAL-AEHKVGASLQHYNPLVDAQTHKTWNLPESWKLRAQMPF 176

Query: 182 GSIEAETGEKEFMNDDDRFKVIG 204
           G+I A  GEK F+ + +RFKV G
Sbjct: 177 GAIAAPAGEKAFIAESERFKVFG 199
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 755

A DNA sequence (GBSx0802) was identified in *S.agalactiae* <SEQ ID 2321> which encodes the amino acid sequence <SEQ ID 2322>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2730(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB62846 GB:AL035475 hypothetical protein [Plasmodium falciparum]
            (ver 2)
 Identities = 112/529 (21%), Positives = 217/529 (40%), Gaps = 67/529 (12%)

Query: 3    NKKHKLLKNIEEFKTITQKRLTERGKFPYDTVHSTFEIKDENFIMERLKSSGLSMGKP-- 60
            N K+  +K  +   ++ Q  + E+ KF  D  H   E + E FI E  +   +   K
Sbjct: 1063 NVKYNEMKGAKN-DSLNQNEIIEKEKF--DLQH---ENRSERFIEEEKQICIVDDKKNNI 1116

Query: 61   --VDYMGVNGIPIYTKTLSIVNKFAFENNSKDSSYSSNINISEDKIKENDQKILDLIVKS 118
              VD    +  P Y + L +       +N  +  YS+      DKI +N++   ++  K
Sbjct: 1117 MNVDEKRKSDHPSYERVLKMEG-----SNKNEEGYSNT-----DKILKNEKNEKNVNEKK 1166

Query: 119  GANNQNLTDEEKVIAFTKYIGEITNYDNEAYRARNVDTEYYRASDLFSVTERKLAMCVGY 178
            G N++   +E+K     K + E  + ++E       D      +  F        +C
Sbjct: 1167 GENDEKNENEKKEENDEKNVNEKKDENDEKNENEKKDENDNNNNSYFYNNSDTFELCTNS 1226

Query: 179  SVTAARAFNIMGIPSYVVSGKSPQGISHAAVRAYYNRSWHIIDITASTYWKNGNYKTTYS 238
             +      N + IPS     ++ +GI  +      N S   I+       KN N ++  YS
Sbjct: 1227 LIFINNKKNSILIPS-----ENEKGIIGSQKEEEQNISPVKINNKKKDLCKNIN-ESDYS 1280

Query: 239  DFIKEYCIDGYD--VYDPAKTNNRFK-VKYMESNEAFENWIHNNGSKSML-------FIN 288
            D      ++  +  +Y    +N++ + ++  + NE + +  + N S++ L       ++
Sbjct: 1281 DKQYSVLLNSIEKKIYKKCSSNSKIRGIEKKKINEDYVDLKNINCSRNTLEFFLTKKYLK 1340

Query: 289  ESAALKDKKPKDDFVPVTEKEKNELIDKYKKLLSQIPENTQNPGEKNIRDYLKNEYEEIL 348
             S  + ++    +   V EK+K +   K KKL  +I  N   P +  I + + +EY  +
Sbjct: 1341 SSELIINEHDCQNINNVYEKKKKKEQAK-KKLNRKI--NVNIPNDSIIEENMSSEYNFVK 1397

Query: 349  KKDN----LFEHEHAE-------FKESLNLNESFYLQLKKEE-------MKPSDNLKKEE 390
            KK+N     FE + ++        F    N   +  L   +E+       ++  +N K+ E
Sbjct: 1398 KKNNNCMVKFETKRSKSILSSEIFAVKKNKKRATNLMRSEEQFISSIGLVEKGENKKRIE 1457

Query: 391  KPRENSVKERETPAENNDFVSVTEKNNLIDKYKELLSKIPENTQNPGEKNIRN--YLEKE 448
            +  E  +KE+    + N+F     KNNL ++   L  K  EN    G  N     ++++
Sbjct: 1458 EKDEEYIKEK-IKNKKNEF-----KNNLTEQL--LFFKSAENINTSGSFNTEKIRHVKRT 1509

Query: 449  YEELLQKDKLFKHEYTEFTKSLNLNETFYSQLKEGEMKLSENPEKGETN 497
              ++    +    +  ++   K L   E    ++ E + ++++N EKGE N
Sbjct: 1510 KRKVNLSNNFILNNFSNILKKLQRMEEDKIKMDEQKKEINKNNEKGEFN 1558
```

There is also homology to SEQ ID 598.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 756

A DNA sequence (GBSx0803) was identified in *S.agalactiae* <SEQ ID 2323> which encodes the amino acid sequence <SEQ ID 2324>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1243(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 757

A DNA sequence (GBSx0804) was identified in *S.agalactiae* <SEQ ID 2325> which encodes the amino acid sequence <SEQ ID 2326>. This protein is predicted to be 2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate al. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1057(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD35160 GB:AE001693 2-dehydro-3-deoxyphosphogluconate
           aldolase/4-hydroxy-2-oxoglutarate aldolase [Thermotoga maritima]
 Identities = 78/192 (40%), Positives = 118/192 (60%), Gaps = 6/192 (3%)

Query: 14  KIVAVIRGNSQEEAFQAAQACIKGGISAIEIAYTNSKASQVIEQLVTQYTNQEQVVVGAG 73
            KIVAV+R NS EEA + A A  +GG+  IEI +T   A  VI++L   +  ++  ++GAG
Sbjct: 11  KIVAVLRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKEL--SFLKEKGAIIGAG 68

Query: 74  TVLDSETARMAILAGAKFIVSPAFNLQTAKLCNRYAIPYLPGCMTLSEVTTALEAGCEII 133
            TV   E  R A+ +GA+FIVSP  + + ++ C    + Y+PG MT +E+  A++ G  I+
Sbjct: 69  TVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGHTIL 128

Query: 134 KIFPGGTLGTSFISSLKAPLPQVQIMVTGGVNLTNAKDWFLSGVTAIGIGGEFNKLAALG 193
            K+FPG  +G  F+ ++K P P V+ + TGGVNL N  +WF +GV A+G+G    K    G
Sbjct: 129 KLFPGEVVGPQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVK----G 184

Query: 194 EFDKITEMAKQY 205
              D++ E AK +
Sbjct: 185 TPDEVREKAKAF 196
```

There is also homology to SEQ ID 1252.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 758

A DNA sequence (GBSx0805) was identified in *S.agalactiae* <SEQ ID 2327> which encodes the amino acid sequence <SEQ ID 2328>. This protein is predicted to be 2-keto-3-deoxygluconate kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4213(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD35161 GB:AE001693 2-keto-3-deoxygluconate kinase [Thermotoga maritima]
 Identities = 94/329 (28%), Positives = 169/329 (50%), Gaps = 7/329 (2%)

Query: 3   KILFFGEPLIRITPKENDYFADSISTKLFYGGSEVNTARALQGFGQDTKLLSALPNNPIG 62
           K++ FGE ++R++P ++     + S  + YGG+E N A  L   G D   ++ LPNNP+G
Sbjct: 2   KVVTFGEIMLRLSPPDHKRIFQTDSFDVTYGGAEANVAAFLAQMGLDAYFVTKLPNNPLG 61

Query: 63  NSFLQFLKAQGIDTHSIQWVGERVGLYFLEDSFACRKGEVVYDRDHSSLHDFRINQIDFD 122
           ++    L+  G+ T  I   G R+G+YFLE   + R  +VVYDR HS++ + +    D++
Sbjct: 62  DAAAGHLRKFGVKTDYIARGGNRIGIYFLEIGASQRPSKVVYDRAHSAISEAKREDFDWE 121

Query: 123 QLFEGVSLFHFSGITLSLDESIQEITLLLLKEAKKREITISLDLNFRSKLISPKNAKILF 182
           ++ +G   FHFSGIT  L + +  I    LK A ++ +T+S DLN+R++L + + A+ +
Sbjct: 122 KILDGARWFHFSGITPPLGKELPLILEDALKVANEKGVTVSCDLNYRARLWTKEEAQKVM 181

Query: 183 SQFATFADICFG----IEPLMVDSQDTTFFNRDEATIEDVKERMISLINHFDFQVIFHTK 238
             F  + D+       IE ++  S +       +   E   +    +   ++F+ +  T
Sbjct: 182 IPFMEYVDVLIANEEDIEKVLGISVEGLDLKTGKLNREAYAKIAEEVTRKYNFKTVGITL 241

Query: 239 RLQDEWGRNHYQAYI-ANRKQEFVTSKEITTAVNQRIGSGDAFVAGALYQLLQHSDSKTV 297
           R       N++   +  N +  F    EI   +  R+G+GD+F    +Y  L   DS+
Sbjct: 242 RESISATVNYWSVMVFENGQPHFSNRYEI--HIVDRVGAGDSFAGALIYGSLMGFDSQKK 299

Query: 298 IDFAVASASLKCALEGDNMFETVTAVNKV 326
            +FA A++ LK  + GD +  ++  + K+
Sbjct: 300 AEFAAAASCLKHTIPGDFVVLSIEEIEKL 328
```

There is also homology to SEQ ID 1264.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 759

A DNA sequence (GBSx0806) was identified in *S.agalactiae* <SEQ ID 2329> which encodes the amino acid sequence <SEQ ID 2330>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.22    Transmembrane   53 -  69 (  53 -  70)

----- Final Results -----
               bacterial membrane --- Certainty=0.1086(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD36157 GB:AE001768 sugar-phosphate isomerase [Thermotoga maritima]
 Identities = 41/125 (32%), Positives = 61/125 (48%), Gaps = 10/125 (8%)

Query: 1   MKIALINENSQASKNTIIYKELKAVSDEKGFEVFNYGMYGKEEESQLTYVQNGLLTAILL 60
           MKIA+ ++++       + +++K     KG EV ++G Y +E      Y +  ++ +IL
Sbjct: 1   MKIAIASDHAAFE----LKEKVKNYLLGKGIEVEDHGTYSEESVDYPDYAKK-VVQSILS 55
```

```
Query: 61  NSGAADFVITGCGTGIGAMLACNSFPGVVCGFAADPVDAYLFSQVNGGNALSLPFAKGFG 120
           N   ADF I  CGTG+G  +A N + G+       P  A L    N  N L LP   G
Sbjct: 56  NE--ADFGILLCGTGLGMSIAANRYRGIRAALCLFPDMARLARSHNNANILVLP---GRL 110

Query: 121 WGAEL 125
            GAEL
Sbjct: 111 IGAEL 115
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2331> which encodes the amino acid sequence <SEQ ID 2332>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2599(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 159/212 (75%), Positives = 186/212 (87%)

Query: 1   MKIALINENSQASKNTIIYKELKAVSDEKGFEVFNYGMYGKEEESQLTYVQNGLLTAILL 60
           MKIALINENSQA+KN IIY  L  V+D+ G++VFNYGMYG E ESQLTYVQNGLL +ILL
Sbjct: 1   MKIALINENSQAAKNGIIYDALTTVTDKHGYQVFNYGMYGTEGESQLTYVQNGLLASILL 60

Query: 61  NSGAADFVITGCGTGIGAMLACNSFPGVVCGFAADPVDAYLFSQVNGGNALSLPFAKGFG 120
            + AADFV+TGCGTG+GAMLA NSFPGV CGFA++P +AYLFSQ+NGGNALS+PFAKGFG
Sbjct: 61  TTKAADFVVTGCGTGVGAMLALNSFPGVTCGFASEPTEAYLFSQINGGNALSIPFAKGFG 120

Query: 121 WGAELNLRYLFERLFEDEKGGGYPKERAVPEQRNARILSEIKQITYRDLLSVLKEIDQDF 180
           WGAELNL  +FERLF +  GGGYPKERA+PEQRNARILS++K+ITYRDLL+++K+IDQDF
Sbjct: 121 WGAELNLTLIFERLFAEPMGGGYPKERAIPEQRNARILSDLKKITYRDLLAIVKDIDQDF 180

Query: 181 LKETISGEHFQEYFFANCQNQNIADYLKSVLD 212
           LKETISG HFQEYFFAN +   +  YLKSVL+
Sbjct: 181 LKETISGAHFQEYFFANAEPSELVTYLKSVLE 212
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 760

A DNA sequence (GBSx0807) was identified in *S.agalactiae* <SEQ ID 2333> which encodes the amino acid sequence <SEQ ID 2334>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane   10 -  26 (   8 -  26)

----- Final Results -----
               bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 761

A DNA sequence (GBSx0808) was identified in *S.agalactiae* <SEQ ID 2335> which encodes the amino acid sequence <SEQ ID 2336>. This protein is predicted to be gluconate 5-dehydrogenase (fabG). Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1117(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC77223 GB:AE000497 5-keto-D-gluconate 5-reductase [Escherichia
           coli K12]
 Identities = 116/260 (44%), Positives = 165/260 (62%), Gaps = 6/260 (2%)

Query: 6   LKDNFSLEGKVALITGASYGIGFSIATAFARAGATIVFNDIKQELVDKGISAYKKLGIKA 65
           + D FSL GK  LITG++ GIGF +AT   + GA I+ NDI  E  +  +    + GI+A
Sbjct: 1   MNDLFSLAGKNILITGSAQGIGFLLATGLGKYGAQIIINDITAERAELAVEKLHQEGIQA 60

Query: 66  HGYVCDVTDEDGINEMVDKISQDVGVIDILVNNAGIIKRTPMLEMSAADFRQVIDIDLNA 125
                +VT +  I+  V+ I +D+G ID+LVNNAGI +R P  E    ++  VI ++  A
Sbjct: 61  VAAPFNVTHKHEIDAAVEHIEKDIGPIDVLVNNAGIQRRHPFTEFPEQEWNDVIAVNQTA 120

Query: 126 PFIVSKAVLPGMIQKGHGKIINICSMMSELGRETVAAYAAAKGGLKMLTKNIASEYGSAN 185
            F+VS+AV   M+++  GK+INICSM SELGR+T+  YAA+KG +KMLT+ +  E    N
Sbjct: 121 VFLVSQAVTRHMVERKAGKVINICSMQSELGRDTITPYAASKGAVKMLTRGMCVELARHN 180

Query: 186 IQCNGIGPGYIATPQTAPLRERQDDGSRHPFDQFIIAKTPAARWGEAEDLGAPAIFLASD 245
           IQ NGI PGY  T  T  L E +       F  ++  +TPAARWG+ ++L   A+FL+S
Sbjct: 181 IQVNGIAPGYFKTEMTKALVEDE------AFTAWLCKRTPAARWGDPQELIGAAVFLSSK 234

Query: 246 ASNFINGHILYVDGGILAYI 265
           AS+F+NGH+L+VDGG+L  +
Sbjct: 235 ASDFVNGHLLFVDGGMLVAV 254
```

There is also homology to SEQ ID 1242:

```
Identities = 225/264 (85%), Positives = 246/264 (92%)

Query: 6   LKDNFSLEGKVALITGASYGIGFSIATAFARAGATIVFNDIKQELVDKGISAYKKLGIKA 65
           +++ FSL+GK+ALITGASYGIGF IA A+A+AGATIVFNDIKQELVDKG++AY++LGI+A
Sbjct: 1   MENMFSLQGKIALITGASYGIGFEIAKAYAQAGATIVFNDIKQELVDKGLAAYRELGIEA 60

Query: 66  HGYVCDVTDEDGINEMVDKISQDVGVIDILVNNAGIIKRTPMLEMSAADFRQVIDIDLNA 125
           HGYVCDVTDE GI +MV +I  +VG IDILVNNAGII+RTPMLEM+A DFRQVIDIDLNA
Sbjct: 61  HGYVCDVTDEAGIQQMVSQIEDEVGAIDILVNNAGIIRRTPMLEMAAEDFRQVIDIDLNA 120

Query: 126 PFIVSKAVLPGMIQKGHGKIINICSMMSELGRETVAAYAAAKGGLKMLTKNIASEYGSAN 185
           PFIVSKAVLP MI KGHGKIINICSMMSELGRETV+AYAAAKGGLKMLTKNIASE+G AN
Sbjct: 121 PFIVSKAVLPSMIAKGHGKIINICSMMSELGRETVSAYAAAKGGLKMLTKNIASEFGEAN 180

Query: 186 IQCNGIGPGYIATPQTAPLRERQDDGSRHPFDQFIIAKTPAARWGEAEDLGAPAIFLASD 245
           IQCNGIGPGYIATPQTAPLRERQ DGSRHPFDQFIIAKTPAARWG  EDL  PA+FLASD
Sbjct: 181 IQCNGIGPGYIATPQTAPLRERQADGSRHPFDQFIIAKTPAARWGTTEDLAGPAVFLASD 240

Query: 246 ASNFINGHILYVDGGILAYIGKQP 269
           ASNF+NGHILYVDGGILAYIGKQP
Sbjct: 241 ASNFVNGHILYVDGGILAYIGKQP 264
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 762

A DNA sequence (GBSx0809) was identified in *S.agalactiae* <SEQ ID 2337> which encodes the amino acid sequence <SEQ ID 2338>. This protein is predicted to be mannose-specific phosphotransferase system component IIAB. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0886(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD46485 GB:AF130465 mannose-specific phosphotransferase system
           component IIAB [Streptococcus salivarius]
 Identities = 43/107 (40%), Positives = 61/107 (56%), Gaps = 3/107 (2%)

Query: 2   IKIIIVAHGNFPDGILSSLELIAGHQEYVVGINFIAGMSSNDVRVALQREVIDFK---EI 58
            I III +HG F +GI  S  +I G QE V  + F+     +D+       +  F    EI
Sbjct: 3   IGIIIASHGKFAEGIHQSGSMIFGDQEKVQVVTFMPSEGPDDLYAHFNDAIAQFDADDEI 62

Query: 59  LVLTDLLGGTPFNVSSALSVEYTDKKIKVLSGLNLSMLMEAVLSRTM 105
            LVL DL  G+PFN +S ++ E  D+KI +++GLNL ML++A   R M
Sbjct: 63  LVLADLWSGSPFNQASRIAGENPDRKIAIITGLNLPMLIQAYTERMM 109
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2339> which encodes the amino acid sequence <SEQ ID 2340>. Analysis of this protein sequence reveals the following:

```
     Possible site: 41
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF81086 GB:AF228498 AgaF [Escherichia coli]
 Identities = 48/127 (37%), Positives = 71/127 (55%), Gaps = 6/127 (4%)

Query: 1   MIAIIVMGHGHFASGIVSALELIAGKQEKVTAIDFTTEMTAADVQDQLSRALIP---EEE 57
            M++II+ GHG FASG+  A++ I G+Q +  AID     + A +  QL  A+     E+
Sbjct: 1   MLSIILTGHGGFASGMEKAMKQILGEQSQFIAIDVPETSSTALLTSQLEEAIAQLDCEDG 60

Query: 58  TLVLCDLLGGTPFKVAATLMESLPNTTCNVLSGLNLAMLIEASFARQTAASFDDLVSGLI 117
             + L DLLGGTPF+VA+TL    P   C V++G NL +L+E    R+  +  +  V  L
Sbjct: 61  IVFLTDLLGGTPFRVASTLAMQKPG--CEVITGTNLQLLEMVLEREGLSGEEFRVQAL- 117

Query: 118 TCSKEGI 124
             C   G+
Sbjct: 118 ECGHRGL 124
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 73/146 (50%), Positives = 94/146 (64%), Gaps = 3/146 (2%)

Query: 1   MIKIIIVAHGNFPDGILSSLELIAGHQEYVVGINFIAGMSSNDVRVALQREVIDFKEILV 60
```

```
           MI II++ HG+F  GI+S+LELIAG QE V  I+F   M++ DV+  L R +I  +E LV
Sbjct: 1   MIAIIVMGHGHFASGIVSALELIAGKQEKVTAIDFTTEMTAADVQDQLSRALIPEEETLV 60

Query: 61  LTDLLGGTPFNVSSALSVEYTDKKIKVLSGLNLSMLMEAVLSRTMFEHVDDLVDKVITSS 120
           L DLLGGTPF V++ L     +    VLSGLNL+ML+EA  +R      DDLV  +IT S
Sbjct: 61  LCDLLGGTPFKVAATLMESLPNTTCNVLSGLNLAMLIEASFARQTAASFDDLVSGLITCS 120

Query: 121 HEGIVDFSTCLATQTAEATFE--GGI 144
            EGIVD+ T L+ Q   AT +  GGI
Sbjct: 121 KEGIVDWKT-LSQQEDGATDDELGGI 145
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 763

A DNA sequence (GBSx0811) was identified in *S.agalactiae* <SEQ ID 2341> which encodes the amino acid sequence <SEQ ID 2342>. This protein is predicted to be unsaturated glucuronyl hydrolase. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.11   Transmembrane  172 - 188 ( 172 - 188)

----- Final Results -----
               bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB05773 GB:AP001514 unsaturated glucuronyl hydrolase [Bacillus halodurans]
 Identities = 156/370 (42%), Positives = 219/370 (59%), Gaps = 3/370 (0%)

Query: 30  EEAIEKALKQLYINIDYFGEEYPTPATFNNIYKVMDNTEWTNGFWTGCLWLAYEYNQDKK 89
           ++A+    ++   NI   F   +P  +     Y++ +N EWTNGFW+G LWL YEY  D
Sbjct: 4   KQAMTDVAEKTLTNIKRFNGRFPHVSEDGEHYELNNNNEWTNGFWSGILWLCYEYTNDPA 63

Query: 90  LKNIAHKNVLSFLNRINNRIALDHHDLGFLYTPSCTAEYRINGDVKALEATIKAADKLME 149
            +  A   V SF  R+   + LDHHD+GFLY+ S  A++ I  D +A + TI+AAD LM+
Sbjct: 64  FRQAAASTVRSFQQRMEQNLELDHHDIGFLYSLSSKAQWIIERDERAKQLTIEAADVLMK 123

Query: 150 RYQEKGGFIQAWGELG-YKEHYRLIIDCLLNIQLLFFAYEQTGDEKYRQVAVNHFYASAN 208
           R++EK    QAWG  G      R+I+DCL+N+ LLF+A E TG+  YR+ A+ H   +
Sbjct: 124 RWREKIELFQAWGPEGDLSNGGRIIVDCLMNLPLLFWASEVTGNPDYREAAIIHADKTRR 183

Query: 209 NVVRDDSSAFHTFYFDPETGEPLKGVTRQGYSDESSWARGQAWGIYGIPLSYRKMKDYQQ 268
            +VR D S +HTFYF+ ETGE L+G T QGY D S+W+RGQAW IYG  ++YR   + +
Sbjct: 184 FIVRGDDSTYHTFYFNQETGEALRGGTHQGYEDGSTWSRGQAWAIYGFAIAYRYTGNERY 243

Query: 269 IILFKGMTNYFLNRLPEDKVSYWDLIFTDGSGQPRDTSATATAVCGIHEMLKYLPEVDPD 328
           +   K    YF+  LP D V+YWD          RD+SA+A A CGI E+L +L E DPD
Sbjct: 244 LETAKRTAKYFIENLPADYVAYWDFNAPITPDTKRDSSASAIASCGILELLSHLQETDPD 303

Query: 329 KETYKYAMHTMLRSLIEQYSNNELIAGRPLLLHGVYSWHSGKGVDEGNIWGDYYYLEALI 388
           K  ++ ++   + SL+E Y++ +   G  L+  G YS   G   D+  IWGDY+Y EAL+
Sbjct: 304 KAFFQQSVQKQMTSLVENYASEKDAQG--LIKRGSYSVRIGHAPDDYVIWGDYFYTEALM 361

Query: 389 RFYKDWELYW 398
           R  K    YW
Sbjct: 362 RLEKLRNGYW 371
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2343> which encodes the amino acid sequence <SEQ ID 2344>. Analysis of this protein sequence reveals the following:

```
     Possible site: 33
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane  173 - 189 ( 173 - 189)

----- Final Results -----
                 bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 273/395 (69%), Positives = 336/395 (84%)

Query: 4   IKPVKVESIENPKRFLNSRLLTKIEVEEAIEKALKQLYINIDYFGEEYPTPATFNNIYKV 63
           +K + +E I+ P+RF     L++ ++ +A++ ALKQ+ +N+DYF E++PTPAT +N Y +
Sbjct: 5   LKTIALEPIKQPERFTKEDFLSQEDITQALDLALKQVRLNMDYFKEDFPTPATKDNQYAI 64

Query: 64  MDNTEWTNGFWTGCLWLAYEYNQDKKLKNIAHKNVLSFLNRINNRIALDHHDLGFLYTPS 123
           MDNTEWTN FWTGCLWLAYEY+ D  +K +A  N LSFL+R+   I LDHHDLGFLYTPS
Sbjct: 65  MDNTEWTNAFWTGCLWLAYEYSGDDAIKALAQANDLSFLDRVTRDIELDHHDLGFLYTPS 124

Query: 124 CTAEYRINGDVKALEATIKAADKLMERYQEKGGFIQAWGELGYKEHYRLIIDCLLNIQLL 183
           C AE+++    ++ EA +KAADKL++RYQ+KGGFIQAWGELG KE YRLIIDCLLNIQLL
Sbjct: 125 CMAEWKLLKTPESREAALKAADKLVQRYQDKGGFIQAWGELGKKEDYRLIIDCLLNIQLL 184

Query: 184 FFAYEQTGDEKYRQVAVNHFYASANNVVRDDSSAFHTFYFDPETGEPLKGVTRQGYSDES 243
           FFA ++TGD +YR +A+NHFYASAN+V+RDD+SA+HTFYFDPETG+P+KGVTRQGYSD+S
Sbjct: 185 FFASQETGDNRYRDMAINHFYASANHVIRDDASAYHTFYFDPETGDPVKGVTRQGYSDDS 244

Query: 244 SWARGQAWGIYGIPLSYRKMKDYQQIILFKGMTNYFLNRLPEDKVSYWDLIFTDGSGQPR 303
           +WARGQAWGIYGIPL+YR +K+ + I LFKGMT+YFLNRLP+D+VSYWDLIF DGS Q R
Sbjct: 245 AWARGQAWGIYGIPLTYRFLKEPELIQLFKGMTHYFLNRLPKDQVSYWDLIFGDGSEQSR 304

Query: 304 DTSATATAVCGIHEMLKYLPEVDPDKETYKYAMHTMLRSLIEQYSNNELIAGRPLLLHGV 363
           D+SATA AVCGIHEMLK LP+ DPDK+TY+ AMH+MLR+LI+ Y+N +L  G PLLLHGV
Sbjct: 305 DSSATAIAVCGIHEMLKTLPDHDPDKKTYEAAMHSMLRALIKDYANKDLKPGAPLLLHGV 364

Query: 364 YSWHSGKGVDEGNIWGDYYYLEALIRFYKDWELYW 398
           YSWHSGKGVDEGNIWGDYYYLEAL+RFYKDW  YW
Sbjct: 365 YSWHSGKGVDEGNIWGDYYYLEALLRFYKDWNPYW 399
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 764

A DNA sequence (GBSx0812) was identified in *S.agalactiae* <SEQ ID 2345> which encodes the amino acid sequence <SEQ ID 2346>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3035(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC44679 GB:U65015 PTS permease for mannose subunit IIIMan C
          terminal domain [Vibrio furnissii]
 Identities = 63/125 (50%), Positives = 89/125 (70%), Gaps = 1/125 (0%)

Query: 5   PNIVMTRVDERLIHGQ-GQLWVKFLSCNTVIVANDDVSKDHLQQTLMKTVVPESIALRFF 63
           PNIV++R+DERL+HGQ G  WV F   N V+VAND+V+ D +QQ LM+ V+ + IA+RF+
```

```
Sbjct: 2   PNIVLSRIDERLVHGQVGVQWVGFADANIVVVANDEVAADTIQQNLMEMVLADGIAIRFW 61

Query: 64  DIQKVIDIIHKANPAQTIFIIVKDLKDVYRLVAGGVPIKEINIGNIHNGEGKEQVSRSIF 123
            +QK ID IHKA+  Q I ++ K   D  RLV GGVPI  IN+GN+H  +GK Q+S+++
Sbjct: 62  TVQKTIDTIHKASDRQRILLVCKTPHDFRRLVEGGVPIAAINVGNMHYIDGKTQISKTVS 121

Query: 124 LGMKD 128
           +  +D
Sbjct: 122 VDAED 126
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2347> which encodes the amino acid sequence <SEQ ID 2348>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2511(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAA84216 GB:AB019619 unsaturated glucuronyl hydrolase [Bacillus
           sp. GL1]
 Identities = 161/369 (43%), Positives = 220/369 (58%), Gaps = 1/369 (0%)

Query: 32  QALDLALKQVRLNMDYFKEDFPTPATKDNQYAIMDNTEWTNAFWTGCLWLAYEYSGDDAI 91
           QA+  AL     N+  F + FP  +   N+Y + DNT+WT+ FW+G LWL YEY+GD+
Sbjct: 4   QAIGDALGITARNLKKFGDRFPHVSDGSNKYVLNDNTDWTDGFWSGILWLCYEYTGDEQY 63

Query: 92  KALAQANDLSFLDRVTRDIELDHHDLGFLYTPSCMAEWKLLKTPESREAALKAADKLVQR 151
           +  A     SF +R+ R   LDHHD+GFLY+ S  A+W + K   +R+ AL AAD L++R
Sbjct: 64  REGAVRTVASFRERLDRFENLDHHDIGFLYSLSAKAQWIVEKDESARKLALDAADVLMRR 123

Query: 152 YQDKGGFIQAWGELGKKEDY-RLIIDCLLNIQLLFFASQETGDNRYRDMAINHFYASANH 210
           ++   G IQAWG  G  E+  R+IIDCLLN+ LL +A ++TGD  YR +A  H   S
Sbjct: 124 WRADAGIIQAWGPKGDPENGGRIIIDCLLNLPLLLWAGEQTGDPEYRRVAEAHALKSRRF 183

Query: 211 VIRDDASAYHTFYFDPETGDPVKGVTRQGYSDDSAWARGQAWGIYGIPLTYRFLKEPELI 270
           ++R D S+YHTFYFDPE G+ ++G T QG +D S W RGQAWGIYG  L  R+L   +L+
Sbjct: 184 LVRGDDSSYHTFYFDPENGNAIRGGTHQGNTDGSTWTRGQAWGIYGFALNSRYLGNADLL 243

Query: 271 QLFKGMTHYFLNRLPKDQVSYWDLIFGDGSEQSRDSSATAIAVCGIHEMLKTLPDHDPDK 330
           +  K M  +FL R+P+D V YWD          RDSSA+AI  CG+ E+   L + DP++
Sbjct: 244 ETAKRMARHFLARVPEDGVVYWDFEVPQEPSSYRDSSASAITACGLLEIASQLDESDPER 303

Query: 331 KTYEAAMHSMLRALIKDYANKDLKPGAPLLLHGVYSWHSGKGVDEGNIWGDYYYLEALLR 390
           + +  A  + + AL   YA +D       +  G Y    G   D+  IWGDYYYLEALLR
Sbjct: 304 QRFIDAAKTTVTALRDGYAERDDGEAEGFIRRGSYHVRGGISPDDYTIWGDYYYLEALLR 363

Query: 391 FYKDWNPYW 399
             +    YW
Sbjct: 364 LERGVTGYW 372
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 112/160 (70%), Positives = 132/160 (82%), Gaps = 1/160 (0%)

Query: 5   PNIVMTRVDERLIHGQGQLWVKFLSCNTVIVANDDVSKDHLQQTLMKTVVPESIALRFFD 64
           PNI+MTRVDERLIHGQGQLWVKFL+CNTVIVAND VS+D +QQ+LMKTV+P SIA+RFF
Sbjct: 4   PNIIMTRVDERLIHGQGQLWVKFLNCNTVIVANDAVSEDKIQQSLMKTVIPSSIAIRFFS 63

Query: 65  IQKVIDIIHKANPAQTIFIIVKDLKDVYRLVAGGVPIKEINIGNIHNGEGKEQVSRSIFL 124
           IQKVIDIIHKA+PAQ+IFI+VKDL+D   LV GGVPI EINIGNIH  + K  +++ I L
Sbjct: 64  IQKVIDIIHKASPAQSIFIVVKDLQDAKLLVEGGVPITEINIGNIHKTDDKVAITQFISL 123
```

```
Query: 125 GMKDKEIIRKLNQEYHIAFNTKTTPTGNDGAVEVNILDYI 164
           G  DK  IR L  ++H+ FNTKTTP GN  A +V+ILDYI
Sbjct: 124 GETDKSAIRCLAHDHHVVFNTKTTPAGN-SASDVDILDYI 162
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 765

A DNA sequence (GBSx0813) was identified in *S.agalactiae* <SEQ ID 2349> which encodes the amino acid sequence <SEQ ID 2350>. This protein is predicted to be AgaW (agaC). Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.95   Transmembrane  251 - 267 ( 244 - 269)
   INTEGRAL    Likelihood = -4.30   Transmembrane  213 - 229 ( 208 - 230)
   INTEGRAL    Likelihood = -2.71   Transmembrane  149 - 165 ( 148 - 165)
   INTEGRAL    Likelihood = -1.81   Transmembrane   31 -  47 (  31 -  49)
   INTEGRAL    Likelihood = -1.49   Transmembrane  173 - 189 ( 173 - 189)

----- Final Results -----
               bacterial membrane --- Certainty=0.3781(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF81084 GB:AF228498 AgaW [Escherichia coli]
 Identities = 93/295 (31%), Positives = 140/295 (46%), Gaps = 48/295 (16%)

Query: 1   MDISILQAVLIGLWTAFCFSGMLLGL-YTNRCIVLSLGVGVILGDIQTALAVGAISELAY 59
           M+IS+LQA  +G+        M  GL + +R +VL   VG++LGD+ T +  G   EL +
Sbjct: 1   MEISLLQAFALGIIAFIAGLDMFNGLTHMHRPVVLGPLVGLVLGDLHTGILTGGTLELVW 60

Query: 60  MGFGVGAGGTVPPNPIGPGIFGTLMAITTAGTKGKITPEAALALSTPIAVGIQFLQTATY 119
           MG    AG   PPN I   I GT  AITT      + P+ A+ ++ P AV +Q   T  +
Sbjct: 61  MGLAPLAGAQ-PPNVIIGTIVGTAFAITTG-----VKPDVAVGVAVPFAVAVQMGITFLF 114

Query: 120 TAFAGAPETAKK--------ALQAGNFRGFKIAANGT-IWAFAGLGFGLGVLGALSTQTL 170
           +  +G      +        AL A N+       N   + AF  + FG     A   +T+
Sbjct: 115 SVMSGVMSRCARMPRTPILAALNACNYLALLALGNFYFLCAFLPIYFG-----AEHAKTI 169

Query: 171 TDLFALIPPVLLNGLTLAGKMLPAIGFAMILSVMAKKELIPYILLGYVLAVYFGLPVLTP 230
            D+   +P  L++GL +AG ++PAIGFA++L +M K   IPY +LG+V A +  LPVL
Sbjct: 170 IDV---LPQRLIDGLGVAGGIMPAIGFAVLLKIMMKNVYIPYFILGFVAAAWLKLPVL-- 224

Query: 231 TANGDGVLTSVATNSVLGVPTIGVAIIATIFALLDIFRKPAAPTKETKTEGDNQD 285
                                  +A  A   AL+D+ RK   PT+    + + +D
Sbjct: 225 ----------------------AIACPALAMALIDLLRKSPEPTQPAAQKEEFED 257
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2351> which encodes the amino acid sequence <SEQ ID 2352>. Analysis of this protein sequence reveals the following:

```
     Possible site: 52
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -6.37   Transmembrane  220 - 236 ( 214 - 241)
   INTEGRAL    Likelihood = -5.10   Transmembrane  146 - 162 ( 144 - 165)
   INTEGRAL    Likelihood = -1.59   Transmembrane  184 - 200 ( 184 - 202)

----- Final Results -----
               bacterial membrane --- Certainty=0.3548(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC44680 GB:U65015 PTS permease for mannose subunit IIPMan
          [Vibrio furnissii]
 Identities = 86/255 (33%), Positives = 137/255 (53%), Gaps = 11/255 (4%)

Query: 1   MDINLLQALLIGLWTAFCFSGMLLGI-YTNRCIILSFGVGIILGDLPTALSMGAISELAY 59
           M+I L QAL++GL        +  G+ + +R ++L   VG+ILGDL T + +G   EL +
Sbjct: 1   MEIGLFQALMLGLLAFLAGLDLFNGLTHFHRPVVLGPLVGLILGDLHTGILVGGTLELIW 60

Query: 60  MGFGVGAGGTVPPNPIGPGIFGTLMAITSAGKVTPEAALALSTPIAVAIQFLQTFAYTAF 119
           MG    AG   PPN I   I GT  AIT+   V P  A+ ++ P AVA+Q   T  ++A
Sbjct: 61  MGLAPLAGAQ-PPNVIIGTIVGTTFAITT--NVEPNVAVGVAVPFAVAVQMGITLLFSAM 117

Query: 120 AGAPETAKKQLQKGNIRGFK---FAANGTIWAFAFIGLGLGLLGALSMDTLLHLVDYIPP 176
           +       +  +  + RG +   + A   + +F F+   L +   L  D    +V  +P
Sbjct: 118 SAVMSKCDEYAKNADTRGIERVNYFALAVLGSFYFLCAFLPIY--LGADHAGAMVAALPK 175

Query: 177 VLLNGLTVAGKMLPAIGFAMILSVMAKKELIPFVLIGYVCAAYLQIPTIGIAIIGIIFAL 236
            L++GL VAG ++PAIGFA+++ +M K   IP+ ++G+V AA+LQ+P + I       A+
Sbjct: 176 ALIDGLGVAGGIMPAIGFAVLMKIMMKNAYIPYFILGFVAAAWLQLPILAIRCAATAMAI 235

Query: 237 NEFYNK--PKQVDAT 249
            +F  K  P  V+A+
Sbjct: 236 IDFMRKSEPTPVNAS 250
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 203/288 (70%), Positives = 225/288 (77%), Gaps = 28/288 (9%)

Query: 1   MDISILQAVLIGLWTAFCFSGMLLGLYTNRCIVLSLGVGVILGDIQTALAVGAISELAYM 60
           MDI++LQA+LIGLWTAFCFSGMLLG+YTNRCI+LS GVG+ILGD+ TAL++GAISELAYM
Sbjct: 1   MDINLLQALLIGLWTAFCFSGMLLGIYTNRCIILSFGVGIILGDLPTALSMGAISELAYM 60

Query: 61  GFGVGAGGTVPPNPIGPGIFGTLMAITTAGTKGKITPEAALALSTPIAVGIQFLQTATYT 120
           GFGVGAGGTVPPNPIGPGIFGTLMAIT+AG   K+TPEAALALSTPIAV IQFLQT  YT
Sbjct: 61  GFGVGAGGTVPPNPIGPGIFGTLMAITSAG---KVTPEAALALSTPIAVAIQFLQTFAYT 117

Query: 121 AFAGAPETAKKALQAGNFRGFKIAANGTIWAFAGLGFGLGVLGALSTQTLTDLFALIPPV 180
           AFAGAPETAKK LQ GN RGFK AANGTIWAFA +G GLG+LGALS  TL  L   IPPV
Sbjct: 118 AFAGAPETAKKQLQKGNIRGFKFAANGTIWAFAFIGLGLGLLGALSMDTLLHLVDYIPPV 177

Query: 181 LLNGLTLAGKMLPAIGFAMILSVMAKKELIPYILLGYVLAVYFGLPVLTPTANGDGVLTS 240
           LLNGLT+AGKMLPAIGFAMILSVMAKKELIP++L+GYV A Y
Sbjct: 178 LLNGLTVAGKMLPAIGFAMILSVMAKKELIPFVLIGYVCAAY------------------ 219

Query: 241 VATNSVLGVPTIGVAIIATIFALLDIFRKPAAPTKETKTEGDNQDDWI 288
                 L +PTIG+AII  IFAL + + KP      T  +G  QDDWI
Sbjct: 220 ------LQIPTIGIAIIGIIFALNEFYNKP-KQVDATTVQGGQQDDWI 260
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 766

A DNA sequence (GBSx0814) was identified in *S.agalactiae* <SEQ ID 2353> which encodes the amino acid sequence <SEQ ID 2354>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2442(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 767

A DNA sequence (GBSx0815) was identified in *S.agalactiae* <SEQ ID 2355> which encodes the amino acid sequence <SEQ ID 2356>. This protein is predicted to be PTS permease for mannose subunit IIBMan. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.28   Transmembrane  278 - 294 ( 272 - 294)
   INTEGRAL    Likelihood = -3.45   Transmembrane  155 - 171 ( 155 - 174)
   INTEGRAL    Likelihood = -1.59   Transmembrane  250 - 266 ( 250 - 267)

----- Final Results -----
               bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8657> which encodes amino acid sequence <SEQ ID 8658> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:     -9.70
GvH: Signal Score (-7.5): -6.12
     Possible site: 19
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 3 value:  -8.28 threshold:  0.0
   INTEGRAL    Likelihood = -8.28   Transmembrane  254 - 270 ( 248 - 270)
   INTEGRAL    Likelihood = -3.45   Transmembrane  131 - 147 ( 131 - 150)
   INTEGRAL    Likelihood = -1.59   Transmembrane  226 - 242 ( 226 - 243)
   PERIPHERAL  Likelihood =  0.37     175
 modified ALOM score:   2.16

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA57943 GB:U18997 ORF_o290; Geneplot suggests frameshift
          linking to o267, not found [Escherichia coli]
 Identities = 101/278 (36%), Positives = 164/278 (58%), Gaps = 6/278 (2%)

Query: 17  LRQKETTKMTGSKKLAKSDYTKTALRAFYLQNGFNYSNYQGLGYANVIYPALKKYYGDDK 76
            ++ K+ T   GS+ ++K D T+   R+  LQ  FNY   Q  G+    + P LKK Y DDK
Sbjct: 19  VKMKKRTTAMGSE-ISKKDITRLGFRSSLLQASFNYERMQAGGFTWAMLPILKKIYKDDK 77

Query: 77  KALAGALEENVEFYNTNPHFLPFVTSLHLAMLDNERPEEEIRGIKMALMGPLAGIGDSLS 136
              L+ A+++N+EF NT+P+ + F+  L ++M +     + I+G+K+AL GP+AGIGD++
Sbjct: 78  PGLSAAMKDNLEFINTHPNLVGFLMGLLISMEEKGENRDTIKGLKVALFGPIAGIGDAIF 137

Query: 137 QFCLAPLFSTIAASLATDGLVMGPILFFVAMNTILTGIKLVTGMYGYRLGTSFIDKLSEQ 196
             F L P+ + I +S A+ G ++GPILFF A+  ++  +++     GY +G   IDK+ E
Sbjct: 138 WFTLLPIMAGICSSFASQGNLLGPILFF-AVYLLIFFLRVGWTHVGYSVGVKAIDKVREN 196
```

```
Query: 197 MSVISRAANIVGVTVISSLAATQVKLTIPYTFAPEKVTSTTQKIVTVQGMLDKIAPALLP 256
             +I+R+A I+G+TVI  L A+ V + +  +FA       T  +   Q   DK+ P +LP
Sbjct: 197 SQMIARSATILGITVIGGLIASYVHINVVTSFA----IDNTHSVALQQDFFDKVFPNILP 252

Query: 257 ALYTFLMFYLIKNKKWTTYKLVILTVIIGILGSWLGIL 294
             YT LM+Y ++ KK     L+ +T ++ I+ S  GIL
Sbjct: 253 MAYTLLMYYFLRVKKAHPVLLIGVTFVLSIVCSAFGIL 290
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2357> which encodes the amino acid sequence <SEQ ID 2358>. Analysis of this protein sequence reveals the following:

```
     Possible site: 45
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.49   Transmembrane  276 - 292 ( 270 - 292)
   INTEGRAL    Likelihood = -7.01   Transmembrane  151 - 167 ( 149 - 176)
   INTEGRAL    Likelihood = -3.03   Transmembrane  202 - 218 ( 202 - 220)
   INTEGRAL    Likelihood = -2.13   Transmembrane  249 - 265 ( 248 - 265)

----- Final Results -----
               bacterial membrane --- Certainty=0.4397(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAA57943 GB:U18997 ORF_o290; Geneplot suggests frameshift
           linking to o267, not found [Escherichia coli]
 Identities = 104/285 (36%), Positives = 162/285 (56%), Gaps = 7/285 (2%)

Query: 8   NKSMQQLSKEANKMTGSNKLTKKDYLKTALRAFFLQNGFNYNNYQGIGYANVIYPALKKH 67
           N+S   +  +       ++++KKD  +   R+  LQ  FNY   Q  G+   + P LKK
Sbjct: 13  NRSPLPVKMKKRTTAMGSEISKKDITRLGFRSSLLQASFNYERMQAGGFTWAMLPILKKI 72

Query: 68  FGNDKKGLYQALEDNCEFYNTNPHFLPFITSLHLVMLENNRPEEETRNIKMALMGPLAGI 127
           + +DK GL  A++DN EF NT+P+ + F+  L + M E     +  + +K+AL GP+AGI
Sbjct: 73  YKDDKPGLSAAMKDNLEFINTHPNLVGFLMGLLISMEEKGENRDTIKGLKVALFGPIAGI 132

Query: 128 GDSLSQFCLAPLFSTIAASLASDGLVLGPILFFLAMNIILTAIKIGSGLYGYKVGTSFID 187
           GD++  F L P+ + I +S AS G +LGPILFF A+ +++  +++G    GY VG   ID
Sbjct: 133 GDAIFWFTLLPIMAGICSSFASQGNLLGPILFF-AVYLLIFFLRVGWTHVGYSVGVKAID 191

Query: 188 KLSEQMAVVSRMANIVGVTVIAGLAATSVKITVPITFAAGKVDAANTAQKFVTIQGMLDK 247
           K+ E   +++R A I+G+TVI GL A+ V I V  +FA     +    Q F       DK
Sbjct: 192 KVRENSQMIARSATILGITVIGGLIASYVHINVVTSFAIDNTHSVALQQDF------FDK 245

Query: 248 IAPALLPALFTLLMYYLIKNKKWTTYKLVILTVIIGVIGSWLGIL 292
           + P +LP  +TLLMYY ++ KK     L+ +T ++ ++ S  GIL
Sbjct: 246 VFPNILPMAYTLLMYYFLRVKKAHPVLLIGVTFVLSIVCSAFGIL 290
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 224/288 (77%), Positives = 255/288 (87%), Gaps = 4/288 (1%)

Query: 12  HLLKKLRQ--KETTKMTGSKKLAKSDYTKTALRAFYLQNGFNYSNYQGLGYANVIYPALK 69
           +L K ++Q  KE  KMTGS KL K DY KTALRAF+LQNGFNY+NYQG+GYANVIYPALK
Sbjct: 6   NLNKSMQQLSKEANKMTGSNKLTKKDYLKTALRAFFLQNGFNYNNYQGIGYANVIYPALK 65

Query: 70  KYYGDDKKALAGALEENVEFYNTNPHFLPFVTSLHLAMLDNERPEEEIRGIKMALMGPLA 129
           K++G+DKK L  ALE+N EFYNTNPHFLPF+TSLHL ML+N RPEEE R IKMALMGPLA
Sbjct: 66  KHFGNDKKGLYQALEDNCEFYNTNPHFLPFITSLHLVMLENNRPEEETRNIKMALMGPLA 125

Query: 130 GIGDSLSQFCLAPLFSTIAASLATDGLVMGPILFFVAMNTILTGIKLVTGMYGYRLGTSF 189
           GIGDSLSQFCLAPLFSTIAASLA+DGLV+GPILFF+AMN ILT IK+ +G+YGY++GTSF
Sbjct: 126 GIGDSLSQFCLAPLFSTIAASLASDGLVLGPILFFLAMNIILTAIKIGSGLYGYKVGTSF 185

Query: 190 IDKLSEQMSVISRAANIVGVTVISSLAATQVKLTIPYTFAPEKV--TSTTQKIVTVQGML 247
```

```
           IDKLSEQM+V+SR ANIVGVTVI+ LAAT VK+T+P TFA  KV   +T QK VT+QGML
Sbjct: 186 IDKLSEQMAVVSRMANIVGVTVIAGLAATSVKITVPITFAAGKVDAANTAQKFVTIQGML 245

Query: 248 DKIAPALLPALYTFLMFYLIKNKKWTTYKLVILTVIIGILGSWLGILA 295
           DKIAPALLPAL+T LM+YLIKNKKWTTYKLVILTVIIG++GSWLGILA
Sbjct: 246 DKIAPALLPALFTLLMYYLIKNKKWTTYKLVILTVIIGVIGSWLGILA 293
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 768

A DNA sequence (GBSx0816) was identified in *S.agalactiae* <SEQ ID 2359> which encodes the amino acid sequence <SEQ ID 2360>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.37   Transmembrane  135 - 151 ( 135 - 151)

----- Final Results -----
               bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB01924 GB:Z79691 OrfA [Streptococcus pneumoniae]
 Identities = 76/206 (36%), Positives = 124/206 (59%), Gaps = 1/206 (0%)

Query: 428 SWTYNSYPKCDYCQLTSKDRYHLVEGQLHVQRASDIYYHKRWLLTLPQAITLVIDKVSCP 487
           SW Y  YP   +C    ++  H +EG          Y HKR +L L + + L++D + C
Sbjct: 2   SWEYEYYPHSLFCHHKEREGMHYIEGAYWSAEPDLPYLHKRKILMLVEDVWLLVDDIRCQ 61

Query: 488 GEHVLTNQYILDDQVIYENGFVNDLKLVSPTTFNLEDCLISKRYNQLTESHKLVKKIKFV 547
           G+H    Q+ILD  V Y++G +N L+L S   F+LED +IS +YN+L  S KL K+  F
Sbjct: 62  GQHEALTQFILDKDVTYQDGKINQLRLWSEVDFDLEDTIISPKYNELERSSKLTKRQFFE 121

Query: 548 DEVMDYTLIVDRNCQVKYVPLVQTNSHKELSNSIAFDIRSQDFHYLIGVLMDDIIFGDKL 607
           ++++DYT+I   + ++    + QT+  +E+ N++AF++++ +   LI +L +DI  G+KL
Sbjct: 122 NQMLDYTIIAHESFEIIRHSVYQTDD-REVENALAFEVKNDETDKLILLLSEDIRVGEKL 180

Query: 608 YLMQGIKCKGKVIVYDKNNGKMSRLK 633
            L+ G K +GK +VYDK N +M RL+
Sbjct: 181 CLVDGTKMRGKCLVYDKINERMIRLQ 206
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2361> which encodes the amino acid sequence <SEQ ID 2362>. Analysis of this protein sequence reveals the following:

```
     Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.55   Transmembrane  477 - 493 ( 477 - 493)

----- Final Results -----
               bacterial membrane --- Certainty=0.2020(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB01924 GB:Z79691 OrfA [Streptococcus pneumoniae]
 Identities = 75/207 (36%), Positives = 125/207 (60%), Gaps = 2/207 (0%)

Query: 434 SWAYLSYPKSNYCHLRQNGHVYFIEGSYQTQFSDRNNYQHDRQILILPPGIFLIIDTIQA 493
           SW Y  YP S +CH ++   +++IEG+Y +   D   Y H R+IL+L   ++L++D I+
Sbjct: 2   SWEYEYYPHSLFCHHKEREGMHYIEGAYWSAEPDLP-YLHKRKILMLVEDVWLLVDDIRC 60
```

```
Query: 494 QGNHCLVSQFILDNHLDVKTDHLSDLRLISDCPFTIEETILSKKYNQYLTSHKLIKRKPF 553
           QG H  ++QFILD  +  +   ++ LRL S+  F +E+TI+S KYN+   S KL KR+ F
Sbjct: 61  QGQHEALTQFILDKDVTYQDGKINQLRLWSEVDFDLEDTIISPKYNELERSSKLTKRQFF 120

Query: 554 KDKGCTSTLLVPDDTKVTPLTPLQTGKRNPIETALSWHLKGKQFDYSICVLQEDLIKGEK 613
           +++    T++  +  ++   +  QT  R  +E AL++ +K  + D  I +L ED+  GEK
Sbjct: 121 ENQMLDYTIIAHESFEIIRHSVYQTDDRE-VENALAFEVKNDETDKLILLLSEDIRVGEK 179

Query: 614 LVLLNSHKIRGKVVVINHITNEIIRLK 640
           L L++  K+RGK +V + I   +IRL+
Sbjct: 180 LCLVDGTKMRGKCLVYDKINERMIRLQ 206
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 282/631 (44%), Positives = 414/631 (64%), Gaps = 2/631 (0%)

Query: 6   YNKFKD-FDREFCQKYIKTYQSNAYQEMKASVNLMMRNTFVFNDNWDMEPCSKAYCLDPL 64
           + +FK+  + +FC+ Y+  YQ+++Y + K   +L++ NTF+F DNWDMEPC   Y LDP+
Sbjct: 11  FARFKETVNPDFCRNYLLDYQTDSYADQKRIADLLLTNTFLFEDNWDMEPCHIPYHLDPI 70

Query: 65  EWDKPVTDDPEWLYMLNRQTYLFKFLVVYIVEGDKSYLRQMKYFMYHWIDCQFTLKPEGA 124
            W + V DDPEW +MLNRQTYL K ++VY+VE D+ YL   K F+ +WI+    L P+G
Sbjct: 71  TWQEAVIDDPEWNFMLNRQTYLQKLILVYLVERDERYLLTAKGFILNWIESAIPLDPKGL 130

Query: 125 VSRTIDTGIRCMSWLKVLIFLDYFGLITETKKIKLLTSLREQITYMRDYYREKDSLSNWG 184
            +RT+DTGIRC +W+K LI+L+ F  +T+ ++  +L SL +Q+ ++   Y +K SLSNWG
Sbjct: 131 ATRTLDTGIRCFAWVKCLIYLNLFNALTKQEESLILASLEKQLQFLHANYLDKYSLSNWG 190

Query: 185 ILQTTAILACLYYYEDELNLPEIQSFAEEELLLQIKLQILDDGSQYEQSIMYHVEVLKSL 244
           ILQTTAIL    Y+  +L++    +FA +EL  QI LQIL+DGSQ+EQS MYHVEVLK+L
Sbjct: 191 ILQTTAILLADAYFGSDLDIAAATAFARKELTQQIALQILEDGSQFEQSTMYHVEVLKAL 250

Query: 245 MELVILAPKYYLPLEETIEKMVTYLIAMTGPDYCQLAIGDSDVTDTRDILTLATLVLKSS 304
           +EL  L P Y   L  T+  M  YL+ MTGPD+ Q+ +GDSDVTDTRDILTLA  +L+
Sbjct: 251 LELTALVPDYLPQLRPTLLAMSDYLLKMTGPDHKQIPLGDSDVTDTRDILTLAATILEEP 310

Query: 305 KTKSFSFDNVNLETLLLFGKPSIYLFEEIPRATIGESAYLFPDSGHVCLRDDRRYIFFKN 364
             K+ +F  +++++LLL G+  ++ FE++P  T+   A+ F  SGH+ +  +  Y+FFKN
Sbjct: 311 HLKAAAFPTLDIDSLLLLGEKGVHTFEQLPVQTLPTFAHHFEHSGHITINQENYYLFFKN 370

Query: 365 GPFGSAHTHSDNNSVCLYDKKKPIFIDAGRYTYKEEQLRYDFKRSTSHSTCTLDGQPLEM 424
           GP GS+HTHSD NS+CLY K +P+F DAGRYTYKEE LRY  K ++ HST  L+ Q  E
Sbjct: 371 GPIGSSHTHSDQNSLCLYYKGQPLFCDAGRYTYKEEPLRYALKSASHHSTAFLEEQLPEQ 430

Query: 425 IKDSWTYNSYPKCDYCQLTSKDRYHLVEGQLHVQRAS-DIYYHKRWLLTLPQAITLVIDK 483
           I  SW Y SYPK +YC L      + +EG    Q +  + Y H R +L LP  I L+ID
Sbjct: 431 IDSSWAYLSYPKSNYCHLRQNGHVYFIEGSYQTQFSDRNNYQHDRQILILPPGIFLIIDT 490

Query: 484 VSCPGEHVLTNQYILDDQVIYENGFVNDLKLVSPTTFNLEDCLISKRYNQLTESHKLVKK 543
           +   G H L +Q+ILD+ +  +   ++DL+L+S   F +E+ ++SK+YNQ   SHKL+K+
Sbjct: 491 IQAQGNHCLVSQFILDNHLDVKTDHLSDLRLISDCPFTIEETILSKKYNQYLTSHKLIKR 550

Query: 544 IKFVDEVMDYTLIVDRNCQVKYVPLVQTNSHKELSNSIAFDIRSQDFHYLIGVLMDDIIF 603
             F D+    TL+V  + +V  +  +QT     +  ++++ ++ + F Y I VL +D+I
Sbjct: 551 KPFKDKGCTSTLLVPDDTKVTPLTPLQTGKRNPIETALSWHLKGKQFDYSICVLQEDLIK 610

Query: 604 GDKLYLMQGIKCKGKVIVYDKNNGKMSRLKN 634
           G+KL L+   K +GKV+V +    ++ RLK+
Sbjct: 611 GEKLVLLNSHKIRGKVVVINHITNEIIRLKH 641
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 769

A DNA sequence (GBSx0817) was identified in *S.agalactiae* <SEQ ID 2363> which encodes the amino acid sequence <SEQ ID 2364>. This protein is predicted to be RegR (kdgR). Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2545(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB01925 GB:Z79691 RegR [Streptococcus pneumoniae]
 Identities = 222/333 (66%), Positives = 279/333 (83%)

Query: 1   MSKKMTINDIAQLSKTSKTTVSFFLNQKFEKMSDETRQRIQEVIDETGYRPSTIARSLNS 60
            M KK+TI DIA++++TSKTTVSF+LN K+EKMS ETR++I++VI ET Y+PS +ARSLNS
Sbjct: 1   MEKKLTIKDIAEMAQTSKTTVSFYLNGKYEKMSQETREKIEKVIHETNYKPSIVARSLNS 60

Query: 61  KKTKLLGVLIGDITNTFSNQIVKGIEHITKQKGYQIIVGNSNYDAKSEEDYIENMLNLGV 120
            K+TKL+GVLIGDITN+FSNQIVKGIE I  Q GYQ+++GNSNY  +SE+ YIE+ML LGV
Sbjct: 61  KRTKLIGVLIGDITNSFSNQIVKGIEDIASQNGYQVMIGNSNYSQESEDRYIESMLLLGV 120

Query: 121 DGFIIQPTSNFRKYSRILKEKKKPMVFFDSQLYEHKTSWVKANNYDAVYDMTQECLNRGY 180
            DGFIIQPTSNFRKYSRI+ EKKK MVFFDSQLYEH+TSWVK NNYDAVYDMTQ C+ +GY
Sbjct: 121 DGFIIQPTSNFRKYSRIIDEKKKKMVFFDSQLYEHRTSWVKTNNYDAVYDMTQSCIEKGY 180

Query: 181 KKFIMITADTSLLSTRIERASGFMDALKDNGFGYDTLVIEDDDHSKSDIEDFLKAVVPDK 240
            + F++ITADTS LSTRIERASGF+DAL D    + +L IED   +   I++FL+  +
Sbjct: 181 EYFLLITADTSRLSTRIERASGFVDALTDANMRHASLTIEDKHTNLEQIKEFLQKEIDPD 240

Query: 241 EETLVFAPNCWALPMVFTAMKNLNFDMPRVGLVGFDNIEWTDFSSPKVSTIVQPAYEEGE 300
            E+TLVF PNCWALP+VFT +K LN+++P+VGL+GFDN EWT FSSP VST+VQP++EEG+
Sbjct: 241 EKTLVFIPNCWALPLVFTVIKELNYNLPQVGLIGFDNTEWTCFSSPSVSTLVQPSFEEGQ 300

Query: 301 QVAQILINRIEGDDSVDNQQIVDCQMFWKESTF 333
            Q  +ILI++IEG +  + QQ++DC + WKESTF
Sbjct: 301 QATKILIDQIEGRNQEERQQVLDCSVNWKESTF 333
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2365> which encodes the amino acid sequence <SEQ ID 2366>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2928(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 214/333 (64%), Positives = 266/333 (79%), Gaps = 2/333 (0%)

Query: 1   MSKKMTINDIAQLSKTSKTTVSFFLNQKFEKMSDETRQRIQEVIDETGYRPSTIARSLNS 60
            M +K+TI DIA+L+KTSKTTVSF+LN +F+KMS+ET+ RI E I  T Y+PS  ARSLN+
Sbjct: 13  MQRKVTIKDIAELAKTSKTTVSFYLNGRFDKMSEETKNRISESIKATNYKPSIAARSLNA 72

Query: 61  KKTKLLGVLIGDITNTFSNQIVKGIEHITKQKGYQIIVGNSNYDAKSEEDYIENMLNLGV 120
            K TKL+GV+IGDITN+FSNQIVKGIE   ++ GYQII+GNSNYD   E++ IE MLNLGV
Sbjct: 73  KSTKLIGVVIGDITNSFSNQIVKGIESKAQEFGYQIIIGNSNYDPSREDELIEKMLNLGV 132
```

```
Query: 121 DGFIIQPTSNFRKYSRILKEKKKPMVFFDSQLYEHKTSWVKANNYDAVYDMTQECLNRGY 180
           DGFIIQPTSNFRKYSRI+  KKK +VFFDSQLYEH+T+WVK NNYDAVYD  Q+C+++GY
Sbjct: 133 DGFIIQPTSNFRKYSRIIDIKKKKVVFFDSQLYEHRTNWVKTNNYDAVYDTIQQCIDKGY 192

Query: 181 KKFIMITADTSLLSTRIERASGFMDALKDNGFGYDTLVIEDDDHSKSDIEDFLKAVVPDK 240
           + FIMIT + +LLSTRIERASGF+D L+ N   +  ++I+++  S   I  FL+  +  K
Sbjct: 193 EHFIMITGNPNLLSTRIERASGFIDVLEANHLTHQEMIIDENQTSSEAIAQFLQGSLTKK 252

Query: 241 EETLVFAPNCWALPMVFTAMKNLNFDMPRVGLVGFDNIEWTDFSSPKVSTIVQPAYEEGE 300
             +LVF PNCWALP VFTAMK+L F++P +GLVGFDNIEWT FSSP ++TI+QPAYEEGE
Sbjct: 253 --SLVFVPNCWALPKVFTAMKSLKFNIPEIGLVGFDNIEWTKFSSPTLTTIIQPAYEEGE 310

Query: 301 QVAQILINRIEGDDSVDNQQIVDCQMFWKESTF 333
           Q  +ILI+ IEG      QQI DCQ+ W+ESTF
Sbjct: 311 QATKILIDDIEGHSQEAKQQIFDCQVNWQESTF 343
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 770

A DNA sequence (GBSx0818) was identified in *S.agalactiae* <SEQ ID 2367> which encodes the amino acid sequence <SEQ ID 2368>. This protein is predicted to be polypeptide defromylase (def-1). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2339(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAC15392 GB:AJ278785 polypeptide deformylase [Streptococcus pneumoniae]
 Identities = 169/204 (82%), Positives = 192/204 (93%), Gaps = 1/204 (0%)

Query: 1   MSAIDKLVKASHLIDMNDIIREGNPTLRKVAEEVTFPLSEKEEILGEKMMQFLKHSQDPI 60
           MSAI+++ KA+HLIDMNDIIREGNPTLR +AEEVTFPLS++E ILGEKMMQFLKHSQDP+
Sbjct: 1   MSAIERITKAAHLIDMNDIIREGNPTLRAIAEEVTFPLSDQEIILGEKMMQFLKHSQDPV 60

Query: 61  MAEKLGLRGGVGLAAPQLDISKRIIAVLVPNVEDAQGNPPKEAYSLQEVMYNPKVVSHSV 120
           MAEK+GLRGGVGLAAPQLDISKRIIAVLVPN+ + +G  P+EAY L+ +MYNPK+VSHSV
Sbjct: 61  MAEKMGLRGGVGLAAPQLDISKRIIAVLVPNIVE-EGETPQEAYDLEAIMYNPKIVSHSV 119

Query: 121 QDAALSDGEGCLSVDREVPGYVVRHARVTIEYFDKTGEKHRLKLKGYNSIVVQHEIDHID 180
           QDAAL +GEGCLSVDR VPGYVVRHARVT++YFDK GEKHR+KLKGYNSIVVQHEIDHI+
Sbjct: 120 QDAALGEGEGCLSVDRNVPGYVVRHARVTVDYFDKDGEKHRIKLKGYNSIVVQHEIDHIN 179

Query: 181 GIMFYDRINEKNPFAVKEGLLILE 204
           GIMFYDRINEK+PFAVK+GLLILE
Sbjct: 180 GIMFYDRINEKDPFAVKDGLLILE 203
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2369> which encodes the amino acid sequence <SEQ ID 2370>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1745(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 160/204 (78%), Positives = 186/204 (90%)

Query: 1   MSAIDKLVKASHLIDMNDIIREGNPTLRKVAEEVTFPLSEKEEILGEKMMQFLKHSQDPI 60
            MSA DKL+K SHLI M+DIIREGNPTLR VA+EV+ PL +++ +LGEKMMQFLKHSQDP+
Sbjct: 1   MSAQDKLIKPSHLITMDDIIREGNPTLRAVAKEVSLPLCDEDILLGEKMMQFLKHSQDPV 60

Query: 61  MAEKLGLRGGVGLAAPQLDISKRIIAVLVPNVEDAQGNPPKEAYSLQEVMYNPKVVSHSV 120
            MAEKLGLR GVGLAAPQ+D+SKRIIAVLVPN+ D +GNPPKEAYS QEV+YNPK+VSHSV
Sbjct: 61  MAEKLGLRAGVGLAAPQIDVSKRIIAVLVPNLPDKEGNPPKEAYSWQEVLYNPKIVSHSV 120

Query: 121 QDAALSDGEGCLSVDREVPGYVVRHARVTIEYFDKTGEKHRLKLKGYNSIVVQHEIDHID 180
            QDAALSDGEGCLSVDR V GYVVRHARVT++Y+DK G++HR+KLKGYN+IVVQHEIDHI+
Sbjct: 121 QDAALSDGEGCLSVDRVVEGYVVRHARVTVDYYDKEGQQHRIKLKGYNAIVVQHEIDHIN 180

Query: 181 GIMFYDRINEKNPFAVKEGLLILE 204
            G++FYDRIN KNPF  KE LLIL+
Sbjct: 181 GVLFYDRINAKNPFETKEELLILD 204
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 771

A DNA sequence (GBSx0819) was identified in *S.agalactiae* <SEQ ID 2371> which encodes the amino acid sequence <SEQ ID 2372>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3620(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10177> which encodes amino acid sequence <SEQ ID 10178> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC75224 GB:AE000305 putative transcriptional regulator
           [Escherichia coli K12]
 Identities = 58/191 (30%), Positives = 98/191 (50%)

Query: 37  DLQVITLTAGQSVCKQGEQLEYLHYIVKGRFKIVRRLFNGKEHILDIKTKPTLIGDIELL 96
            D ++    A   + ++G+Q  +L Y+ +GR ++   L NG+  ++D    P  IG+IEL+
Sbjct: 17  DTRLFHFLARDYIVQEGQQPSWLFYLTRGRARLYATLANGRVSLIDFFAAPCFIGEIELI 76

Query: 97  TNRQIVSSVIALEDLTVIQLSLKGRKEKLLTDATFLLKLSQELAQAFHDQNIKASTNLGY 156
                   +V A+E+   + L +K  +  LL D  FL KL   L+   +   +  + N  +
Sbjct: 77  DKDHEPRAVQAIEECWCLALPMKHYRPLLLNDTLFLRKLCVTLSHKNYRNIVSLTQNQSF 136

Query: 157 TVKELLASHILAIEEQGYFQLELSSLADSFGVSYRHLLRVIHDMVKEGLIQKEKPKYFIK 216
             +   LA+ IL  +E   +  + +  A+  GVSYRHLL V+   + +GL+ K K  Y IK
Sbjct: 137 PLVNRLAAFILLSQEGDLYHEKHTQAAEYLGVSYRHLLYVLAQFIHDGLLIKSKKGYLIK 196

Query: 217 NRFALESLNIQ 227
            NR  L  L ++
Sbjct: 197 NRKQLSGLALE 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2373> which encodes the amino acid sequence <SEQ ID 2374>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3809(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 23/63 (36%), Positives = 35/63 (55%), Gaps = 1/63 (1%)

Query: 146 QNIKASTNLGYTVKELLASHILAIEEQGYFQLELSSLADSFGVSYRHLLRVIHDMVKEGL 205
           QN+    N+ YTVKE  AS+ L  +      L L+ LA+ FG S RHL  V+   + + +
Sbjct: 3   QNV-CQQNITYTVKERFASYTLEAQANQEVHLNLTLLANRFGTSDRHLKHVLKQPIFQRI 61

Query: 206 IQK 208
           I++
Sbjct: 62  IER 64
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 772

A DNA sequence (GBSx0820) was identified in *S.agalactiae* <SEQ ID 2375> which encodes the amino acid sequence <SEQ ID 2376>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.24   Transmembrane  163 - 179 ( 159 - 185)
   INTEGRAL    Likelihood = -8.49   Transmembrane  204 - 220 ( 201 - 226)
   INTEGRAL    Likelihood = -7.80   Transmembrane  272 - 288 ( 269 - 296)
   INTEGRAL    Likelihood = -6.00   Transmembrane  333 - 349 ( 331 - 352)
   INTEGRAL    Likelihood = -5.41   Transmembrane   75 -  91 (  73 -  92)
   INTEGRAL    Likelihood = -4.94   Transmembrane  245 - 261 ( 240 - 262)
   INTEGRAL    Likelihood = -4.41   Transmembrane  362 - 378 ( 359 - 380)
   INTEGRAL    Likelihood = -4.14   Transmembrane   96 - 112 (  95 - 113)
   INTEGRAL    Likelihood = -2.44   Transmembrane  141 - 157 ( 141 - 158)
   INTEGRAL    Likelihood = -1.81   Transmembrane  302 - 318 ( 301 - 320)

----- Final Results -----
               bacterial membrane --- Certainty=0.4694(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8659> which encodes amino acid sequence <SEQ ID 8660> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:      -3.52
GvH: Signal Score (-7.5): 0.340001
     Possible site: 25
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 11 value:  -9.24 threshold:  0.0
   INTEGRAL    Likelihood = -9.24   Transmembrane  134 - 150 ( 130 - 156)
   INTEGRAL    Likelihood = -8.60   Transmembrane   17 -  33 (  13 -  37)
   INTEGRAL    Likelihood = -8.49   Transmembrane  175 - 191 ( 172 - 197)
   INTEGRAL    Likelihood = -7.80   Transmembrane  243 - 259 ( 240 - 267)
   INTEGRAL    Likelihood = -6.00   Transmembrane  304 - 320 ( 302 - 323)
   INTEGRAL    Likelihood = -5.41   Transmembrane   46 -  62 (  44 -  63)
   INTEGRAL    Likelihood = -4.94   Transmembrane  216 - 232 ( 211 - 233)
   INTEGRAL    Likelihood = -4.41   Transmembrane  333 - 349 ( 330 - 351)
   INTEGRAL    Likelihood = -4.14   Transmembrane   67 -  83 (  66 -  84)
   INTEGRAL    Likelihood = -2.44   Transmembrane  112 - 128 ( 112 - 129)
```

```
  INTEGRAL    Likelihood = -1.81   Transmembrane  273 - 289 ( 272 - 291)
  PERIPHERAL  Likelihood =  3.45     193
modified ALOM score:   2.35

*** Reasoning Step: 3

----- Final Results -----
                bacterial membrane --- Certainty=0.4694(Affirmative) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB50057 GB:AJ248286 TRANSPORT PROTEIN, permease [Pyrococcus abyssi]
 Identities = 94/382 (24%), Positives = 173/382 (44%), Gaps = 30/382 (7%)

Query: 5   MEKLSLLSL-SLILLSTFSTSPALPQMISYY-RDKGLPSPQVELLFSIPSMAIIFILLIT 62
           MEKL +L L SL  +  +S   A+P +     +D G+ + ++ LL +   +    I +
Sbjct: 1   MEKLIILILISLGWIFNYSHRMAVPSLAPIIMKDLGINNAEIGLLMTSLLLPYSLIQVPA 60

Query: 63  PWLSKKLSEKHMIIFGLLLTALGGGLPVVSQNYLLVFVSRLLLGSGIGFINTRAISVISE 122
            ++  K+  K ++   +L  +L   L V++++Y  +   R L G   G     A ++ISE
Sbjct: 61  GYIGDKIGRKKLLTISILGYSLSSALIVLTRDYWDLVTVRALYGFFAGLYYAPATALISE 120

Query: 123 YYQGKERRKLLGLRGSFEVLGNA---GLTAL--VGLLLTFGWSKSFMIYFLALPILVLYL 177
            ++ ++   L      F ++G A   G+T L  V + LT  W  +F++  +   I+ + L
Sbjct: 121 LFRERKGSAL-----GFFMVGPAIGSGITPLIVVPVALTLSWRYAFLVLSIMSSIVGILL 175

Query: 178 VFAPKKVVKDTNDKIKTKGQKIPKADLTYIVALAILAGFVITINTGINLRIPLLVVEFGL 237
           + A K       + IK +G K       ++++LA   G        +   +  LV   G+
Sbjct: 176 MVAIK------GEPIKVEGVKFKIPRGVFLLSLANFLGLGAFFAM-LTFLVSYLVSR-GV 227

Query: 238 GTPAQASLVLSAMMLMGIIAGMSFGQLIAMFHKQLIPICLVLFS-LTLLGVGLPSNLMVL 296
           G   +ASL+ S + L+GI+  +  G L     K  + +   L S LT L + +PS L ++
Sbjct: 228 GME-KASLMFSMLSLVGILGSIIAGFLYDHLGKVSVLLAYALNSLLTFLVIVIPSPLFLI 286

Query: 297 TISAMASGFLYSL--MVTAVFSLVADRVEYSLVGSATTLVLVF-CNIGGASAAILLSCFD 353
            +  +    LYS+  ++TA  S  A R    +V     +V  F   IG      L+
Sbjct: 287 PLGLV----LYSVGGIMTAYTSEKASRENLGVVMGFVNMVGFFGATIGPYIVGFLIDRLG 342

Query: 354 HLLGQINAVFYVYAILSLAVGM 375
           + L  + +V   Y + ++ +G+
Sbjct: 343 YSLALL-SVPLAYLVSAVIIGL 363
```

There is also homology to SEQ ID 2378.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 773

A DNA sequence (GBSx0821) was identified in *S.agalactiae* <SEQ ID 2379> which encodes the amino acid sequence <SEQ ID 2380>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38   Transmembrane  171 - 187 ( 171 - 187)

----- Final Results -----
              bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB61731 GB:AL133220 putative oxidoreductase. [Streptomyces
            coelicolor A3(2)]
 Identities = 101/327 (30%), Positives = 169/327 (50%), Gaps = 12/327 (3%)

Query: 8    WATLGTGVIANEL-AQALEARGQKLYSVANRTYDKGLEFATKYGIQKVYDHIDQVFEDPE 66
            W  L TG +A    A  ++    ++ +VA+RT      FA ++GI + Y   + +  D +
Sbjct: 11   WGILATGGMAARFTADLVDLPDAEVVAVASRTEASAKTFAERFGIPRAYGGWETLARDED 70

Query: 67   VDIIYISTPHNTHISFLRKALANGKHVLCEKSITLNSTELKEAIDLAETNHVVLAEAMTI 126
            VD++Y++TPH+ H +     L  G++VLCEK  TLN+ E  E + LA  N V L EAM +
Sbjct: 71   VDVVYVATPHSAHRTAAGLCLEAGRNVLCEKPFTLNAREAAELVALARENGVFLMEAMWM 130

Query: 127  FHMPIYRQLKTLVDSGKLGPLKMIQMNFGSYKEYDMTNRFFSRDLAGGALLDIGVYALSC 186
            +  P+ R+LK LV  G +G ++ +Q +FG    +   +R       GGALLD+GVY +S
Sbjct: 131  YCNPLVRRLKELVADGAIGEVRSLQADFGLAGPFPAAHRLRDPAQGGGALLDLGVYPVSF 190

Query: 187  IRWFMSEAPHNITSQVTFAPTGVDEQVGILLTNPANEMATVSLSLHAKQPKRATIAYDKG 246
             +  + E P ++ ++   +  GVD Q G LL+   + +A++  S+    P  A+I   +G
Sbjct: 191  AQLLLGE-PTDVAARAVLSEEGVDLQTGALLSYGNDALASIHCSITGGTPNSASITGSEG 249

Query: 247  YIEL---FEYPRGQKAVITYTEDGHQDIL--EAGKTENALQYEVADMEEAV-SGKTNH-- 298
             I++   F +P     V+  T    Q+     A     +L++E  ++  A+ +G+T
Sbjct: 250  RIDVPNGFFFP--DHFVLHRTGRDPQEFRADPADGPRESLRHEAEEVMRALRAGETESPL 307

Query: 299  MYLNYTKDVMDIMTQLRQEWGFTYPEE 325
            + L+ T  VM  +  +R   G  YP E
Sbjct: 308  VPLDGTLAVMRTLDAIRDRVGVRYPGE 334
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 774

A DNA sequence (GBSx0822) was identified in *S.agalactiae* <SEQ ID 2381> which encodes the amino acid sequence <SEQ ID 2382>. This protein is predicted to be oligopeptidase. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2881(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAC14579 GB:AJ249396 oligopeptidase [Streptococcus thermophilus]
 Identities = 504/631 (79%), Positives = 563/631 (88%)

Query: 1    MIKYQDDFYQAVNGEWAKTAVIPDDKPRTGGFSDLADDIEALMLSTTDKWLADENKPSDT 60
            M + QDDFY A+NGEW KTAVIPDDKP TGGFSDLAD+IE LML TTD+WLA EN P +
Sbjct: 1    MTRLQDDFYHAINGEWEKTAVIPDDKPCTGGFSDLADEIEDLMLETTDQWLAGENVPDNA 60

Query: 61   ILNHFIAFHKMTADYQKREEVGVSPVLPLIEEYKGLQSFSEFASKVAEYELEGKPNEFPF 120
            IL +FI FH+MTADY +RE VG+ PV PLIEEYK L SFSEFASK+AEYE+ GKPNEFPF
Sbjct: 61   ILQNFIKFHRMTADYDRREAVGIEPVKPLIEEYKKLSSFSEFASKIAEYEMSGKPNEFPF 120

Query: 121  GVAPDFMNAQLNVLWAEAPGIILPDTTYYSEDNEKGKELLAFWRKSQEDLLPLFGLSEQE 180
             V+PDFMNAQLNVLWA+APGIILPDTTYY+EDNEKGKELL  WR+ QE+LL  +G + +E
Sbjct: 121  SVSPDFMNAQLNVLWADAPGIILPDTTYYTEDNEKGKELLEIWREMQEELLGKYGFTAEE 180

Query: 181  IKDILDKVLALDAKLAQYVLSREESSEYVKLYHPYNWEDFTKLAPELPLDAIFQKILGQK 240
            IKD+LDKV+ LDAKLA+YVLS EESSEYV+LYHPY+W DFTKLAPELPLD+IF +ILGQ
Sbjct: 181  IKDLLDKVIDLDAKLAKYVLSHEESSEYVELYHPYDWADFTKLAPELPLDSIFTEILGQV 240
```

```
Query: 241 PDKVIVPEERFWTEFASDYYSESNWELLKADLILSAANAYNAYLTDDIRIKSGVYSRALS 300
           PDKVIV EE FWTEFA++YYSE+NWELLKA L++ A  ++NAYLTD++R+ SG YSRALS
Sbjct: 241 PDKVIVSEESFWTEFAAEYYSEANWELLKAVLLIDATTSWNAYLTDELRVLSGKYSRALS 300

Query: 301 GTPQAMDKKKAAYYLASGPYNQALGLWYAGEKFSPEAKADVEHKIATMIDVYKSRLEKAD 360
           GTPQAMDKKKAA+YLA GPYNQALGLWYAGEKFSPEAKADVE K+ATMIDVYKSRL+ AD
Sbjct: 301 GTPQAMDKKKAAFYLAQGPYNQALGLWYAGEKFSPEAKADVEAKVATMIDVYKSRLQTAD 360

Query: 361 WLAQSTREKAIMKLNVITPHIGYPEKLPETYTKKIIDPKLSLVENATNLDKISIAYGWSK 420
           WLA  TREKAI KLNVITPHIGYPEKLPETY KKIID  LSLVENA  L +ISIA+ WSK
Sbjct: 361 WLAPETREKAITKLNVITPHIGYPEKLPETYDKKIIDENLSLVENAQKLVEISIAHSWSK 420

Query: 421 WNKPVDRSEWHMPAHMVNAYYDPQQNQIVFPAAILQEPFYALEQSSSANYGGIGAVIAHE 480
           WNKPVDRSEWHMPAHMVNAYYDPQQNQIVFPAAILQ PFY + QSSSANYGGIGAVIAHE
Sbjct: 421 WNKPVDRSEWHMPAHMVNAYYDPQQNQIVFPAAILQAPFYDIAQSSSANYGGIGAVIAHE 480

Query: 481 ISHAFDTNGASFDEHGSLNNWWTDEDFEAFKKLTDKVVEQFDGLESYGAKVNGKLTVSEN 540
           ISHAFDTNGASFDE+GSL NWWT++D+ AFK+ TDK+V+QF+GL+SYGAKVNGKLTVSEN
Sbjct: 481 ISHAFDTNGASFDENGSLKNWWTEDDYAAFKERTDKIVDQFEGLDSYGAKVNGKLTVSEN 540

Query: 541 VADLGGVACALEAAQRESDFSARDFFINFATIWRMKARDEYMQMLASVDVHAPAQWRTNI 600
           VADLGGVACALEAA+R+ DFS R+FFINFATIWR KAR+EYMQMLASVDVHAPA+WRTN+
Sbjct: 541 VADLGGVACALEAAKRDEDFSVREFFINFATIWRTKAREEYMQMLASVDVHAPAKWRTNV 600

Query: 601 TVTNFEEFHKEFDVKDGDNMWRPVEKRVIIW 631
            VTNF+EFHKEFDVK+GD MWR  E RVIIW
Sbjct: 601 IVTNFDEFHKEFDVKEGDGMWRAPEDRVIIW 631
```

Endopeptidases are often exposed antigens.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2383> which encodes the amino acid sequence <SEQ ID 2384>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2622(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 504/631 (79%), Positives = 564/631 (88%)

Query: 1   MIKYQDDFYQAVNGEWAKTAVIPDDKPRTGGFSDLADDIEALMLSTTDKWLADENKPSDT 60
           M  YQDDFYQAVNG+WA+TAVIPDDKPRTGGFSDLAD+IEALML TTD WLA EN P D
Sbjct: 1   MTTYQDDFYQAVNGKWAETAVIPDDKPRTGGFSDLADEIEALMLDTTDAWLAGENIPDDA 60

Query: 61  ILNHFIAFHKMTADYQKREEVGVSPVLPLIEEYKGLQSFSEFASKVAEYELEGKPNEFPF 120
           IL +F+ FH++ ADY KR+EVGVSP+LPLIEEY+ L+SFSEF + +A+YEL G PNEFPF
Sbjct: 61  ILKNFVKFHRLVADYAKRDEVGVSPILPLIEEYQSLKSFSEFVANIAKYELAGLPNEFPF 120

Query: 121 GVAPDFMNAQLNVLWAEAPGIILPDTTYYSEDNEKGKELLAFWRKSQEDLLPLFGLSEQE 180
            VAPDFMNAQLNVLWAEAP I+LPDTTYY E NEK +EL   WR+SQE LLP FG S +E
Sbjct: 121 SVAPDFMNAQLNVLWAEAPSILLPDTTYYEEGNEKAEELRGIWRQSQEKLLPQFGFSTEE 180

Query: 181 IKDILDKVLALDAKLAQYVLSREESSEYVKLYHPYNWEDFTKLAPELPLDAIFQKILGQK 240
           IKD+LDKV+ LD +LA+YVLSREE SEY KLYHPY W DF KLAPELPLD+IF+KILGQ
Sbjct: 181 IKDLLDKVIELDKQLAKYVLSREEGSEYAKLYHPYVWADFKKLAPELPLDSIFEKILGQV 240

Query: 241 PDKVIVPEERFWTEFASDYYSESNWELLKADLILSAANAYNAYLTDDIRIKSGVYSRALS 300
           PDKVIVPEERFWTEFA+ YYSE+NW+LLKA+LI+ AANAYNAYLTDDIR++SG YSRALS
Sbjct: 241 PDKVIVPEERFWTEFAATYYSEANWDLLKANLIVDAANAYNAYLTDDIRVESGAYSRALS 300

Query: 301 GTPQAMDKKKAAYYLASGPYNQALGLWYAGEKFSPEAKADVEHKIATMIDVYKSRLEKAD 360
```

```
           GTPQAMDK+KAA+YLA GP++QALGLWYAG+KFSPEAKADVE K+A MI+VYKSRLE AD
Sbjct: 301 GTPQAMDKQKAAFYLAQGPFSQALGLWYAGQKFSPEAKADVESKVARMIEVYKSRLETAD 360

Query: 361 WLAQSTREKAIMKLNVITPHIGYPEKLPETYTKKIIDPKLSLVENATNLDKISIAYGWSK 420
           WLA +TREKAI KLNVITPHIGYPEKLPETY KK+ID  LSLVENA NL KI+IA+ WSK
Sbjct: 361 WLAPATREKAITKLNVITPHIGYPEKLPETYAKKVIDESLSLVENAQNLAKITIAHTWSK 420

Query: 421 WNKPVDRSEWHMPAHMVNAYYDPQQNQIVFPAAILQEPFYALEQSSSANYGGIGAVIAHE 480
           WNKPVDRSEWHMPAH+VNAYYD QQNQIVFPAAILQEPFY+L+QSSSANYGGIGAVIAHE
Sbjct: 421 WNKPVDRSEWHMPAHLVNAYYDLQQNQIVFPAAILQEPFYSLDQSSSANYGGIGAVIAHE 480

Query: 481 ISHAFDTNGASFDEHGSLNNWWTDEDFEAFKKLTDKVVEQFDGLESYGAKVNGKLTVSEN 540
           ISHAFDTNGASFDEHGSLN+WWT ED+ AFK+ TDK+V QFDGLES+GAKVNGKLTVSEN
Sbjct: 481 ISHAFDTNGASFDEHGSLNDWWTQEDYAAFKERTDKIVAQFDGLESHGAKVNGKLTVSEN 540

Query: 541 VADLGGVACALEAAQRESDFSARDFFINFATIWRMKARDEYMQMLASVDVHAPAQWRTNI 600
           VADLGGVACALEAAQ E DFSARDFFINFATIWRMKAR+EYMQMLAS+DVHAP + RTN+
Sbjct: 541 VADLGGVACALEAAQSEEDFSARDFFINFATIWRMKAREEYMQMLASIDVHAPGELRTNV 600

Query: 601 TVTNFEEFHKEFDVKDGDNMWRPVEKRVIIW 631
           T+TNF+ FH+ FD+K+GD MWR  + RVIIW
Sbjct: 601 TLTNFDAFHETFDIKEGDAMWRAPKDRVIIW 631
```

SEQ ID 2382 (GBS193) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 3; MW 73kDa).

The GBS193-His fusion product was purified (Figure 196, lane 5) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 253). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 775

A DNA sequence (GBSx0823) was identified in *S.agalactiae* <SEQ ID 2385> which encodes the amino acid sequence <SEQ ID 2386>. This protein is predicted to be immunity protein (mccF-1). Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1627(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9433> which encodes amino acid sequence <SEQ ID 9434> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB84435 GB:AF027868 YocD [Bacillus subtilis]
 Identities = 114/270 (42%), Positives = 170/270 (62%), Gaps = 4/270 (1%)

Query: 1   MSFSKHYLENDILYSASITSRVEDLHEAFADPSVDAILATIGGFNSNELLPYLDYDLISK 60
           ++ ++H  E +   S+SI SRV DLH AF DP V AIL T+GGFNSN+LL YLDY+ I +
Sbjct: 43  VTIAEHANECNEFDSSSIESRVHDLHAAFFDPGVKAILTTLGGFNSNQLLRYLDYEKIKR 102

Query: 61  NPKIICGYSDSTAFLNAIFAKAKIQTYMGPAYSSFKMKEGQPYQTQAWLT-AMTENHYEL 119
           +PKI+CGYSD TA  NAI+ K  + TY GP +S+F MK+G  Y  + +L+   +++ +E+
Sbjct: 103 HPKILCGYSDITALCNAIYQKTGLVTYSGPHFSTFAMKKGLDYTEEYFLSCCASDDPFEI 162
```

```
Query: 120 WPSEEWSSDPWYDPSKPRQFFPTEWK-IYNHGKASGTIIGGNLSTFGLLRGTPYAPKIER 178
            PS EWS D W+   + R+F+P     +   G A GT+IGGNL T  LL+GT Y P+ E
Sbjct: 163 HPSSEWSDDRWFLDQENRRFYPNNGPVVIQEGYAEGTLIGGNLCTLNLLQGTEYFPETEH 222

Query: 179 YVLLIEEAEESNFYEFDRNLAAI--LQAYPHPQAILMGRFPKECGMTPQVFEYILSKHAI 236
            +LLIE+   S+ + FDR+L ++  L A+ H +AIL+GRF K   ++  + + ++
Sbjct: 223 TILLIEDDYMSDIHMFDRDLQSLIHLPAFSHVKAILIGRFQKASNVSIDLVKAMIETKKE 282

Query: 237 FKEIPVIYDMDFAHTQPLLTVTIGAELSVD 266
              IP+I +++  HT P+ T  IG     ++
Sbjct: 283 LSGIPIIANINAGHTSPIATFPIGGTCRIE 312
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2387> which encodes the amino acid sequence <SEQ ID 2388>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1162(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 75/252 (29%), Positives = 125/252 (48%), Gaps = 22/252 (8%)

Query: 34  VDAILATIGGFNSNELLPYLDYDLISKNPKIICGYSDSTAFLNAIFAKAKIQTYMGPAYS 93
            VD I+ +IGG+NSN +L Y+DYDL  +    I  GYSD+TA   A++ K    TY+  +
Sbjct: 1   VDVIMTSIGGYNSNSVLKYIDYDLFKQKFPIFIGYSDTTALALALYKKTGCITYLSQSVI 60

Query: 94  SFKMKEGQP----------YQTQAWLTAMTENHYELWPSEEWSSDPWYDPSKPRQFFPTE 143
            S    E +P          +  Q+    +     ++W ++EW +  W    + ++    E
Sbjct: 61  S-NFGEFEPFNELNYFYFDFMLQSKCETLMVQIPDVW-TDEWIN--WETYERTKKTNKNE 116

Query: 144 WKIYNHGKASGTIIGGNLSTFGLLRGTPYAPKIERYVLLIEEAEESNFYEFDRNLA--AI 201
            W I+N G+ +GT+IGGNL T   + GT Y PKI    +L+ E   ++     RN    A+
Sbjct: 117 WIIFNKGEFNGTLIGGNLDTIVGIIGTEYMPKITEDTILLLEDVYTDLGRLYRNFTTLAL 176

Query: 202 LQAYPHPQAILMGRFPKECGMTPQVFEYILSKHAIFKEIPVIYDMDFAHTQPLLTVTIGA 261
               +     +++ +F +  G    V   I+++    ++IP++ + D  HT P   + IG
Sbjct: 177 HGIFDKIGGLIISKF-ETIGENSDVINDIINEFVGHRKIPILLNFDCGHTHPSCLMPIGG 235

Query: 262 ELSVDTTTLSLS 273
            ++     TLSLS
Sbjct: 236 KI-----TLSLS 242
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 776

A DNA sequence (GBSx0824) was identified in *S.agalactiae* <SEQ ID 2389> which encodes the amino acid sequence <SEQ ID 2390>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3112(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 777

A DNA sequence (GBSx0825) was identified in *S.agalactiae* <SEQ ID 2391> which encodes the amino acid sequence <SEQ ID 2392>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.6171(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10175> which encodes amino acid sequence <SEQ ID 10176> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 778

A DNA sequence (GBSx0826) was identified in *S.agalactiae* <SEQ ID 2393> which encodes the amino acid sequence <SEQ ID 2394>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.19   Transmembrane   83 -  99 (  80 - 113)
   INTEGRAL    Likelihood = -9.71   Transmembrane    4 -  20 (   1 -  24)
   INTEGRAL    Likelihood = -9.45   Transmembrane  315 - 331 ( 307 - 337)
   INTEGRAL    Likelihood = -8.33   Transmembrane  186 - 202 ( 180 - 210)
   INTEGRAL    Likelihood = -7.75   Transmembrane  233 - 249 ( 227 - 255)
   INTEGRAL    Likelihood = -3.98   Transmembrane  390 - 406 ( 382 - 407)
   INTEGRAL    Likelihood = -3.61   Transmembrane   27 -  43 (  27 -  45)
   INTEGRAL    Likelihood = -3.29   Transmembrane  107 - 123 ( 105 - 125)
   INTEGRAL    Likelihood = -1.75   Transmembrane  273 - 289 ( 273 - 290)

----- Final Results -----
               bacterial membrane --- Certainty=0.5076(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15347 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 174/524 (33%), Positives = 275/524 (52%), Gaps = 13/524 (2%)

Query: 1    MEETILIVSFLLFLILSNVINRIFPKLPLPFIQLVFGILSGLVFHKSQVHIDPELFLAFV 60
            M+  ++++  L  + +SN++NR  P +P+P IQ+  GIL+          ++ ELF
Sbjct: 1    MDIFLVVLVLLTIIAISNIVNRFIPFIPVPLIQVALGILAASFPQGLHFELNTELFFVLF 60

Query: 61   IAPLNFREGQESDIGSFIKYRAIILYLILPTVFLTAIVVGYVAGHLLPVSLPLAACFALG 120
            IAPL F +G+ +        RA IL L L  VF T IV GY    ++P ++PLAA F L
Sbjct: 61   IAPLLFNDGKRTPRAELWNLRAPILLLALGLVFATVIVGGYTIHWMIP-AIPLAAAFGLA 119
```

```
Query: 121 AALGPTDAVAFISIAKRFQFPKRAENILKLEGLLNDASGLVSFQFALTALVTGYFSLAKA 180
           A L PTD VA  +++ R + PK    +L+ EGL+NDASGLV+F+FA+ A VTG FSLA+A
Sbjct: 120 AILSPTDVVAVSALSGRVKMPKGILRLLEGEGLMNDASGLVAFKFAIAAAVTGAFSLAQA 179

Query: 181 SLKLALAIMGGFLIGLLFAFLMRLCLTVLEKFDAADVTGALLLELTLPFVAYFVADLLGF 240
           ++      +GG L G++ +FL+      L +    DVT  +L+++  PFV Y  A+ +G
Sbjct: 180 AVSFVFISLGGLLCGVVISFLIIRFRLFLRRLGMQDVTMHMLIQILTPFVIYLAAEEIGV 239

Query: 241 SAIIAVVVAGVMQANRLKKVTLFDAQVDRVTSVIWETLNFILNGLVFLIFGRELTRIIGP 300
           S I+AVV  G+  A    ++     ++  V+S  W  + FILNGLVF+I G ++  +I
Sbjct: 240 SGILAVVAGGITHAVEQDRLESTMIKLQIVSSSTWNIILFILNGLVFVILGTQIPDVISV 299

Query: 301 LLTSNAYSNFDLISIVVLVTCTLFLVRFLAVSCFY--AWRSFKYHKSFKKYWREIQLLTF 358
           +    A SN  +I  ++++T TL L+RFL V  F+   W   K    +K   R   L++
Sbjct: 300 IFNDTAISNMKVIGYILVITFTLMLLRFLWVLFFWNGKWFFNKDQNIYKPGLRSTLLISI 359

Query: 359 SGVKGSVSIATILLLPKHSVIGE--LGYSLILFTVGAVTLMSFLTGLLVLPKLAPPLQVK 416
           SGV+G+V++A    +P     G      +LILF    V L + +   +VLP L    +
Sbjct: 360 SGVRGAVTLAGSFSIPYFLEDGTPFPERNLILFLAAGVILCTLVIATVVLPILTEKEEED 419

Query: 417 DD-----YLIRLSILTKVLSVLEEDGKSSENQASFYAVIDNYNSRIRHLILEQ--ESSDI 469
           ++        R  ++   L  ++ED   +   AS  AVI  YN ++++L  +Q   S+ I
Sbjct: 420 EERNKKLLTARRKLIKTALQTIKEDMNETNKTASL-AVIAEYNEKMKNLRFQQYTSSNRI 478

Query: 470 KKDLAELQLMMLSIESDGLEAAYRYGNISIKEYRIYQRYLKYLE 513
           KK   +++   +  E + L      G+I  +   + Q     LE
Sbjct: 479 KKHERKVRAQGVKAEQEALMKMLERGDIPEETANVLQERFNELE 522
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 779

A DNA sequence (GBSx0827) was identified in *S.agalactiae* <SEQ ID 2395> which encodes the amino acid sequence <SEQ ID 2396>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3494(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 780

A DNA sequence (GBSx0828) was identified in *S.agalactiae* <SEQ ID 2397> which encodes the amino acid sequence <SEQ ID 2398>. This protein is predicted to be integrase (phage-relatedpr). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
                bacterial cytoplasm --- Certainty=0.5094(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10173> which encodes amino acid sequence <SEQ ID 10174> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF12706 GB:AF066865 integrase [bacteriophage TPW22]
 Identities = 171/353 (48%), Positives = 253/353 (71%), Gaps = 1/353 (0%)

Query: 21  MASYRKRENGLWEYRISYKTIDGKYKRKEKGGFKTKKLAQAAAIEIEKKLTQNILTNDEV 80
           MA++RKR    W++R+SYK  +G+YK+ EKGG+KTKK A+AAA E +K+L  +   ++++
Sbjct: 1   MANFRKRGK-TWQFRLSYKDNNGEYKKFEKGGYKTKKEAEAAADEAKKRLNNHSEFDNDI 59

Query: 81  TLYDFVKTWSEVYKRPYVKDKTWETYSKNFKHIKNYFQELKVKDITPLYYQKKLNEFGEK 140
           +LYDF + W++VYK+P+V + TW TY +    I  Y ++  + +ITP +YQ  LN+
Sbjct: 60  SLYDFFEKWAKVYKKPHVTEATWRTYKRTLNLIDKYIKDKPIAEITPTFYQAVLNKMSLL 119

Query: 141 YAQETLEKFHYQIKGAMKVAVREQVVTFNFAEGAKVKSQVEPKNEEEDFLEEREYKALLA 200
           Y QE+L+KF++QIK AMK+AV E+V++ NFA+  K KS++  +  EE +L   EY  LLA
Sbjct: 120 YRQESLDKFYFQIKSAMKIAVHEKVISENFADFTKAKSKLAARPVEEKYLHADEYLKLLA 179

Query: 201 LTRENIQYVSYFTLYLLAVTGLRFSEAMGLTWSDIDFKNGILDINKSFDYSNTQDFADLK 260
           +  E ++Y SYF  YL AVTG+RF+E +GLTWS +DF    + I +++DYS T +FA+ K
Sbjct: 180 IAEEKMEYTSYFACYLTAVTGMRFAELLGLTWSHVDFDKKEISIQRTWDYSITNNFAETK 239

Query: 261 NESSKRKVPIDSNTIDILREYKKNHWQANIKNRVCFGVSNSACNKLIKKIVGRKVRNHSL 320
           NESSKRK+PI S TI +L++YKK +W  N  +RV + +SN+  NK IK I GRKV  HSL
Sbjct: 240 NESSKRKIPISSKTIKLLKKYKKEYWHENKYDRVIYNLSNNGLNKTIKVIAGRKVHPHSL 299

Query: 321 RHTYASFLILNGVDIVTISKLLGHESPDITLKVYTHQMEALAERNFEKIKNIF 373
           RH++AS+LI  G+D++T+SKLLGHE+ ++TLKVY HQ++ + + N + I+ IF
Sbjct: 300 RHSFASYLIYKGIDLLTVSKLLGHENLNVTLKVYAHQLKEMEQENNDVIRKIF 352
```

There is also homology to SEQ ID 578.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 781

A DNA sequence (GBSx0829) was identified in *S.agalactiae* <SEQ ID 2399> which encodes the amino acid sequence <SEQ ID 2400>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3377(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 782

A DNA sequence (GBSx0830) was identified in *S.agalactiae* <SEQ ID 2401> which encodes the amino acid sequence <SEQ ID 2402>. This protein is predicted to be homology to cI-like repressor. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0827(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD44097 GB:AF115103 orf122 gp [Streptococcus thermophilus
           bacteriophage Sfi21]
 Identities = 57/125 (45%), Positives = 77/125 (61%), Gaps = 5/125 (4%)

Query: 3   MKLDQLCKEFGVELCLFDASDWHSSGFYNPITKVLGVDVNLSEQEQKQVALHELQHKNHF 62
           M   +L ++FGV LC F +S W   GF +P+ +V+ ++ +L  + + +V LHEL H  H
Sbjct: 1   MNESELLEQFGVSLCEFSSSQWTRDGFLDPVNRVVYINRDLPTERRLKVLLHELGHLEHD 60

Query: 63  PYQYQLFRERCELDANRNMIHHLLKEELEIAEDHTQFNYLVFMEKYKLKTIADEAMIKEE 122
           P QY+  RE+ E  ANRNMIH LLK      E+   FNY+ FMEKY L TI DE  +K E
Sbjct: 61  PKQYERLREKYEAQANRNMIHELLKN-----ENLDNFNYVHFMEKYNLTTICDETFVKNE 115

Query: 123 YLNLV 127
           YL L+
Sbjct: 116 YLKLI 120
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 783

A DNA sequence (GBSx0831) was identified in *S.agalactiae* <SEQ ID 2403> which encodes the amino acid sequence <SEQ ID 2404>. This protein is predicted to be EpsR protein. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4692(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF12710 GB:AF066865 repressor protein [bacteriophage TPW22]
 Identities = 36/101 (35%), Positives = 62/101 (60%), Gaps = 7/101 (6%)

Query: 4   LIDRIRELSNKKGMSLNDLEDTLGYSRNSLYSLNE-NSKMGKPKEIAQYFNVSLDYLLGL 62
           L ++I+EL+++K +S+  +E+ LG++   ++   + N  + K K++A+YFNVS+D+LLGL
Sbjct: 3   LYEKIKELASQKNVSIRQVEEKLGFANGTIRQWGKKNPGINKVKDVAKYFNVSVDFLLGL 62

Query: 63  TDNPRIAS--DETAIIDGQVVDLREAAAHTMLFDGKPLDED 101
            DN R     D    +D   V+  E +     FDGKPL ++
Sbjct: 63  DDNQRKKEPVDLADFVDDNKVNWDEWVS----FDGKPLSDE 99
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 784

A DNA sequence (GBSx0832) was identified in *S.agalactiae* <SEQ ID 2405> which encodes the amino acid sequence <SEQ ID 2406>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4079(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 785

A DNA sequence (GBSx0833) was identified in *S.agalactiae* <SEQ ID 2407> which encodes the amino acid sequence <SEQ ID 2408>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2942(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10171> which encodes amino acid sequence <SEQ ID 10172> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 786

A DNA sequence (GBSx0834) was identified in *S.agalactiae* <SEQ ID 2409> which encodes the amino acid sequence <SEQ ID 2410>. This protein is predicted to be a replication initiation protein Rep (RC). Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3335(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 787

A DNA sequence (GBSx0835) was identified in *S.agalactiae* <SEQ ID 2411> which encodes the amino acid sequence <SEQ ID 2412>. This protein is predicted to be antirepressor. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3380(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA97816 GB:AB044554 antirepressor [Staphylococcus aureus
           prophage phiPV83]
 Identities = 70/153 (45%), Positives = 93/153 (60%), Gaps = 15/153 (9%)

Query: 3   EIFVFHGQEVRTVTINNEPWFVGKDVADILGYSKSRNAIALHVDEDDALKQGITDNLGRM 62
           + F F    VRTV I NEP+FVGKD+A+ILGY+++ NAI  HVD +D L    + + G+
Sbjct: 5   QTFNFKELPVRTVEIENEPYFVGKDIAEILGYARTDNAIRNHVDSEDKLTHQFSAS-GQN 63

Query: 63  QETIIINESGLYSLIL----SSKLPQVKE----FKRWVTSEVLPQIRQQGAYVPENLSDE 114
           +  IIINESGLYSLI      SK  +++E    FKRWVTS+VLP IR+ G Y  +N+ ++
Sbjct: 64  RNMIIINESGLYSLIFDASKQSKNEKIRETARKFKRWVTSDVLPAIRKHGIYATDNVIEQ 123

Query: 115 A------FIALFTGQKKLKEHQLALAQDVDYLK 141
                   I + T  KK KE  L L Q V+  K
Sbjct: 124 TLKDPDYIITVLTEYKKEKEQNLVLQQQVEVNK 156
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2413> which encodes the amino acid sequence <SEQ ID 2414>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4609(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 54/142 (38%), Positives = 73/142 (51%), Gaps = 7/142 (4%)

Query: 11  EVRTVTINNEPWFVGKDVADILGYSKSRNAIALHVDEDDALKQGITDNLGRMQETIIINE 70
           EVRT TINN+ +F   D   IL  S  R  I   +++D      I D+LGR Q+   INE
Sbjct: 13  EVRTATINNQIYFNLNDCCQILELSNPRKTIE-RLNKDGVTTSDIIDSLGRTQQANFINE 71

Query: 71  SGLYSLILSSKLPQVKEFKRWVTSEVLPQIRQQGAYVPENLSDEA------FIALFTGQK 124
           S  Y L+  S+ P+ ++F  WVTSEVLP IR+ GAY+ E   ++A       I L    K
Sbjct: 72  SNFYKLVFQSRKPEAEKFADWVTSEVLPSIRKHGAYMTEQTLEQALTSPDFLIRLANELK 131

Query: 125 KLKEHQLALAQDVDYLKNEQPI 146
           + KE    L  +   L  E  +
```

```
Sbjct: 132 EEKERSRQLEAEKSILSVENMV 153
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 788

A DNA sequence (GBSx0836) was identified in *S.agalactiae* <SEQ ID 2415> which encodes the amino acid sequence <SEQ ID 2416>. This protein is predicted to be e11. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3281(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC27227 GB:AF009630 e11 [bacteriophage bIL170]
 Identities = 66/161 (40%), Positives = 93/161 (56%), Gaps = 13/161 (8%)

Query: 15  YQVSNLGRVRSIGRTVNAKQRTRKTKGRILKQSL-SSGYAIVTLSVNGLRKSIRVHRLVA 73
            Y+VSNLG+VR+I              GRILK  +  +GY +  L  N  +K++ +HR++A
Sbjct: 16  YEVSNLGKVRNI------------KSGRILKPWIVPNGYLMHQLCENNKKKNLFLHRIIA 63

Query: 74  EAFIPNPINKRTINHIDENKLNNRVDNLEWATDKENANHGNRTTKSSLGRCKPVEQFTLE 133
             AFI NP  K  +NHIDENKLNN ++NLEW T KEN  HG R  + +    K V Q  L
Sbjct: 64  TAFIDNPEEKPQVNHIDENKLNNDLNNLEWCTVKENNIHGTRMKRIAEKHFKKVIQLDLN 123

Query: 134 GEFINTFDSIKSASMKTGISSQRITATAMGHQKQTHGYKWR 174
               +N F+S+  A  +TG+S + I++   G +K    +KWR
Sbjct: 124 DNVLNEFESMVQAEQETGVSRRNISSCCNGKRKSAGRFKWR 164
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 789

A DNA sequence (GBSx0837) was identified in *S.agalactiae* <SEQ ID 2417> which encodes the amino acid sequence <SEQ ID 2418>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2357(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10169> which encodes amino acid sequence <SEQ ID 10170> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 790

A DNA sequence (GBSx0838) was identified in *S.agalactiae* <SEQ ID 2419> which encodes the amino acid sequence <SEQ ID 2420>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.47    Transmembrane   21 -  37 (  19 -  38)

----- Final Results -----
               bacterial membrane --- Certainty=0.3187(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 791

A DNA sequence (GBSx0839) was identified in *S.agalactiae* <SEQ ID 2421> which encodes the amino acid sequence <SEQ ID 2422>. This protein is predicted to be DNA polymerase III delta prime subunit (dnaB). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0544(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
AAF98347 AF280763 DNA polymerase III delta prime subunit [Streptococcus pyogenes]
 Identities = 284/444 (63%), Positives = 357/444 (79%), Gaps = 4/444 (0%)

Query: 3   ELKVLPHDIQAEQSVLGSIFIKPEKMIEVAEYLKPNDFYRPAHKILFKAMVSLADRGEAI 62
           EL+V P D+ AEQSVLGSIFI P+K+I V E++ P+DFY+ AHKI+F+AM++L+DR +AI
Sbjct: 8   ELRVQPQDLLAEQSVLGSIFISPDKLIAVREFISPDDFYKYAHKIIFRAMITLSDRNDAI 67

Query: 63  DIVTIKSTLESTDELGMVGGISYIAEIVNAVPTSSHAEHYAKIVAKKAQLRSIIDNLSDS 122
           D  TI++ L+  D+L  +GG+SYI E+VN+VPTS++AE+YAKIVA+KA LR II  L++S
Sbjct: 68  DATTIRTILDDQDDLQSIGGLSYIVELVNSVPTSANAEYYAKIVAEKAMLRDIIARLTES 127

Query: 123 IGNAYDEDMDIDEIIAKAERSLIEVSQASNKSSFRPIHDVLLENHSKIEERSNNTSQITG 182
           +  AYDE +  +E+IA  ER+LIE+++ SN+S FR I DVL  N+  +E RS  TS +TG
Sbjct: 128 VNLAYDEILKPEEVIAGVERALIELNEHSNRSGFRKISDVLKVNYEALEARSKQTSNVTG 187

Query: 183 IETGFYDFDKLITGLHEDQLIVLAARPAMGKTALALNIAQNVATKSNKAVAVFSLEMGAE 242
           + TGF D DK+ TGLH DQL++LAARPA+GKTA  LNIAQNV TK  K VA+FSLEMGAE
Sbjct: 188 LPTGFRDLDKITTGLHPDQLVILAARPAVGKTAFVLNIAQNVGTKQKKTVAIFSLEMGAE 247

Query: 243 SLVERMLSAEGTIINHHIRTGNLTVNEWQRLIYAQGQLAEAPIFIDDTAGVKITDIRARA 302
           SLV+RML+AEG + +H +RTG LT  +W  +  AQG LAEAPI+IDDT G+KIT+IRAR+
Sbjct: 248 SLVDRMLAAEGMVDSHSLRTGQLTDQDWNNVTIAQGALAEAPIYIDDTPGIKITEIRARS 307
```

```
Query: 303 RRLSQETD-GLGLIVIDYLQLIQGSRSDNRQQEVSEISRQLKIIAKELKVPVIALSQLSR 361
           R+LSQE D GLGLIVIDYLQLI G++ +NRQQEVS+ISRQLKI+AKELKVPVIALSQLSR
Sbjct: 308 RKLSQEVDGGLGLIVIDYLQLITGTKPENRQQEVSDISRQLKILAKELKVPVIALSQLSR 367

Query: 362 GVEQRNDKRPIMSDLRESGSIEQDADIVAFLYRDAYYQ---DKKEGQPENDITELIIRKN 418
           GVEQR DKRP++SD+RESGSIEQDADIVAFLYRD YY+   D  E   E++  E+I+ KN
Sbjct: 368 GVEQRQDKRPVLSDIRESGSIEQDADIVAFLYRDDYYRKECDDAEEAVEDNTIEVILEKN 427

Query: 419 RHGNLGTVKLYFHKEYTKFSSVEE 442
           R G  GTVKL F KEY KFSS+ +
Sbjct: 428 RAGARGTVKLMFQKEYNKFSSIAQ 451
```

There is also homology to SEQ ID 2424:

```
 Identities = 284/444 (63%), Positives = 357/444 (79%), Gaps = 4/444 (0%)

Query: 3   ELKVLPHDIQAEQSVLGSIFIKPEKMIEVAEYLKPNDFYRPAHKILFKAMVSLADRGEAI 62
           EL+V P D+ AEQSVLGSIFI P+K+I V E++ P+DFY+ AHKI+F+AM++L+DR +AI
Sbjct: 11  ELRVQPQDLLAEQSVLGSIFISPDKLIAVREFISPDDFYKYAHKIIFRAMITLSDRNDAI 70

Query: 63  DIVTIKSTLESTDELGMVGGISYIAEIVNAVPTSSHAEHYAKIVAKKAQLRSIIDNLSDS 122
           D  TI++ L+  D+L  +GG+SYI E+VN+VPTS++AE+YAKIVA+KA LR II  L++S
Sbjct: 71  DATTIRTILDDQDDLQSIGGLSYIVELVNSVPTSANAEYYAKIVAEKAMLRDIIARLTES 130

Query: 123 IGNAYDEDMDIDEIIAKAERSLIEVSQASNKSSFRPIHDVLLENHSKIEERSNNTSQITG 182
           +  AYDE +  +E+IA  ER+LIE+++ SN+S FR I DVL  N+  +E RS  TS +TG
Sbjct: 131 VNLAYDEILKPEEVIAGVERALIELNEHSNRSGFRKISDVLKVNYEALEARSKQTSNVTG 190

Query: 183 IETGFYDFDKLITGLHEDQLIVLAARPAMGKTALALNIAQNVATKSNKAVAVFSLEMGAE 242
           + TGF D DK+ TGLH DQL++LAARPA+GKTA  LNIAQNV TK  K VA+FSLEMGAE
Sbjct: 191 LPTGFRDLDKITTGLHPDQLVILAARPAVGKTAFVLNIAQNVGTKQKKTVAIFSLEMGAE 250

Query: 243 SLVERMLSAEGTIINHHIRTGNLTVNEWQRLIYAQGQLAEAPIFIDDTAGVKITDIRARA 302
           SLV+RML+AEG + +H +RTG LT  +W  +  AQG LAEAPI+IDDT G+KIT+IRAR+
Sbjct: 251 SLVDRMLAAEGMVDSHSLRTGQLTDQDWNNVTIAQGALAEAPIYIDDTPGIKITEIRARS 310

Query: 303 RRLSQETD-GLGLIVIDYLQLIQGSRSDNRQQEVSEISRQLKIIAKELKVPVIALSQLSR 361
           R+LSQE D GLGLIVIDYLQLI G++ +NRQQEVS+ISRQLKI+AKELKVPVIALSQLSR
Sbjct: 311 RKLSQEVDGGLGLIVIDYLQLITGTKPENRQQEVSDISRQLKILAKELKVPVIALSQLSR 370

Query: 362 GVEQRNDKRPIMSDLRESGSIEQDADIVAFLYRDAYYQ---DKKEGQPENDITELIIRKN 418
           GVEQR DKRP++SD+RESGSIEQDADIVAFLYRD YY+   D  E   E++  E+I+ KN
Sbjct: 371 GVEQRQDKRPVLSDIRESGSIEQDADIVAFLYRDDYYRKECDDAEEAVEDNTIEVILEKN 430

Query: 419 RHGNLGTVKLYFHKEYTKFSSVEE 442
           R G  GTVKL F KEY KFSS+ +
Sbjct: 431 RAGARGTVKLMFQKEYNKFSSIAQ 454
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 792

A DNA sequence (GBSx0840) was identified in *S.agalactiae* <SEQ ID 2425> which encodes the amino acid sequence <SEQ ID 2426>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2146(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10167> which encodes amino acid sequence <SEQ ID 10168> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 793

A DNA sequence (GBSx0841) was identified in *S.agalactiae* <SEQ ID 2427> which encodes the amino acid sequence <SEQ ID 2428>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2774(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 794

A DNA sequence (GBSx0842) was identified in *S.agalactiae* <SEQ ID 2429> which encodes the amino acid sequence <SEQ ID 2430>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.91   Transmembrane   63 -  79 (  62 -  79)

----- Final Results -----
               bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8661> which encodes amino acid sequence <SEQ ID 8662> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:    -11.31
GvH: Signal Score (-7.5): -1.86
     Possible site: 28
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -1.91 threshold:  0.0
   INTEGRAL    Likelihood = -1.91   Transmembrane   61 -  77 (  60 -  77)
   PERIPHERAL  Likelihood =  9.92      19
 modified ALOM score:   0.88

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB18686 GB:U38906 ORF11 [Bacteriophage r1t]
 Identities = 101/249 (40%), Positives = 157/249 (62%), Gaps = 21/249 (8%)

Query: 3   MAQRRMFSRKITETDRFLEMPLSSQALYFHLNMGADDEGFIDKAKTIQRTIGASDDDMKL 62
            MAQRRM  ++  +T +FL +PL +QALYFHL + ADD+G ++ A  + R +GA++D + L
Sbjct: 1   MAQRRMIDKRTIQTQKFLRLPLETQALYFHLMLNADDDGVVE-AFPVVRMVGAAEDSLGL 59

Query: 63  LIAKGFLIPFDSGVV-VIRHWRIHNYIQSDRFQSTLYQSEKAQLEYDKSKTASLKPIGNC 121
            L+ K F+ P +  +V  I  ++  N I+ DR++++ Y    AQL  ++      ++P  N
Sbjct: 60  LVVKQFIKPLNEEMVYFIIDFKEQNTIKKDRYKASKY----AQLLTNEEFGTEMEPKRNQ 115

Query: 122 IQNVSKMETQVRLSKGSLDKDSLTTYPTVSDNEEEDIPYKEIISYLNEKANRNYRPNIQK 181
            +    K     RL K  LDK++      +S   ++ IPY EI+ YLN+K  R++R N++
Sbjct: 116 LGTSDKN----RLDKNRLDKNN-----NMSGKPDDVIPYSEILEYLNKKTGRSFR-NVEA 165

Query: 182 NKTLIKARWSEGFRLDDFKHVIDTTVKDWSGTKY-----EKYLRPETLFGSKFEGYLNQA 236
            NK LIKARW+EG++L+DFK V+D  V +WSG  +     E YL+P+TLF +KF+ YLNQ
Sbjct: 166 NKKLIKARWNEGYKLEDFKTVVDNMVSNWSGKMFNGVPAENYLQPKTLFSNKFDSYLNQV 225

Query: 237 PRIKTETID 245
            PRI+ + I+
Sbjct: 226 PRIEQKEIN 234
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8662 (GBS344) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 12; MW 30.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 3; MW 59kDa).

The GBS344-GST fusion product was purified (Figure 213, (lane 3; Figure 226, lanes 4-6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 271), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 795

A DNA sequence (GBSx0843) was identified in *S.agalactiae* <SEQ ID 2431> which encodes the amino acid sequence <SEQ ID 2432>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2549(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG31329 GB:AF182207 ORF 272 [Bacteriophage mv4]
 Identities = 70/241 (29%), Positives = 125/241 (51%), Gaps = 30/241 (12%)

Query: 12  VLEETCEVHGCQLWLTKVPIKGRLEELKQCPECTKAAINIFENKLNSQSKINSKLADTYA 71
            VLE+ C  HG  L +T    +G  E++  CP+C   A+ + + + + +++  S +A
Sbjct: 16  VLEQKCSKHGLNL-ITYKNHEG--EQVTCCPQCQAEALEVLQERFDQKAR-QSIIARK-- 69
```

```
Query: 72  VFERDSLVSDKLRAKSLENYE---------IKDEIDQHAINYAKRMEQFYRQDRTGNAII 122
            F  +SL + K+   + + +E         IK ++   A+ +A +    +       A++
Sbjct: 70  -FRENSLANSKMWKCTFDTFEAQPGSAEELIKGQVRNAAVAFATKPVAHH-------AVL 121

Query: 123 TGPSGVGKSHLTYGLAKFMNEQFKAYESPKSVLFISLVSLFTKIKESFKVDNGY-RQADM 181
            G  G GKSHL    A  M ++   +   K++ FI++  LF+KIK SF   + Y  +
Sbjct: 122 YGQPGAGKSHL----AMAMMQEIHKHRPTKTMAFINISRLFSKIKNSFDDPSEYWTKEKA 177

Query: 182 IELLTRVDYLFLDDLGKESRKGDS--QNNEWTHQILYEILDNRSNTIINTNLSSKEIKALY 240
           +E++  VD L +DDLG ES  G +  +  +W   ++Y++L+N+   II TNLS +E+K +Y
Sbjct: 178 LEIMRGVDLLCIDDLGTESSMGRTGQEATKWAQDVIYDVLENQDRIIITTNLSERELKRVY 238
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 796

A DNA sequence (GBSx0844) was identified in *S.agalactiae* <SEQ ID 2433> which encodes the amino acid sequence <SEQ ID 2434>. This protein is predicted to be methyl transferase. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1241(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10165> which encodes amino acid sequence <SEQ ID 10166> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC98421 GB:L29323 methyl transferase [Streptococcus pneumoniae]
 Identities = 262/474 (55%), Positives = 313/474 (65%), Gaps = 71/474 (14%)

Query: 2   MKFLDLFAGIGGFRLGMEQAGHECIGFCEINKFARASYKVIHDTEGEIELHDITRVSD-E 60
           M+F+DLF+GIGGFRLGME  GHECIGFCEI+KFAR SYK I  TEGEIE HDI  VSD E
Sbjct: 1   MRFIDLFSGIGGFRLGMESVGHECIGFCEIDKFARESYKSIFQTEGEIEFHDIRDVSDDE 60

Query: 61  FIRGIGSVDVICGGFPCQAFSIAGNRRGFEDTRGTLFFEIARFASILRPKYLFLENVKGL 120
           F +  G VDVICGGFPCQAFSIAG R GFEDTRGTLFFEIAR A  ++P++LFLENVKGL
Sbjct: 61  FKKLRGKVDVICGGFPCQAFSIAGRRLGFEDTRGTLFFEIARAAKQIQPRFLFLENVKGL 120

Query: 121 LNHEGGATFETIIRTLDELGYNVEWQIFNSKNFGVPQNRERVFIIGHLRGEGTRPIFPFE 180
           LNH+ G TF TI+ TLDELG++VEWQ+ NSK+FGVPQNRERVFIIGH R  GTR  FPF
Sbjct: 121 LNHDKGRTFTTILTTLDELGFDVEWQMLNSKDFGVPQNRERVFIIGHSRKRGTRLGFPFR 180

Query: 181 SSITENYPIHTRKIGNVNPSGNGMNGEVYDSEGLSPTLTTNKGEGVKIAVN--------- 231
                  P   + +GN+NPS +GM+G+VY SEGL+PTL   KGEG KIA+
Sbjct: 181 REGQATNPETLKILGNLNPSKSGMSGKVYYSEGLAPTLVRGKGEGFKIAIPCMTPDRLDK 240

Query: 232 --------------------------VVGRLPGKFEMPNRVYDPDGLAPTIRTMQGGGLE 265
                                     VVG LP  F+   RVY  +GL+PT+ TMQGG
Sbjct: 241 RQNGRRFKDNQEPMFTLNTQDRHGIVVVGDLPTSFKETGRVYGSEGLSPTLTTMQGGDKI 300

Query: 266 PKIIQRGRGYNQGGEYEISPTVTCNSWQENNLLKIKEATKKGYSEAEAGDSVNLSHPNSE 325
           PKI+                       +    LK++EATKKGY++AE GDS+NL  P+S+
Sbjct: 301 PKILIP---------------------EPIQFLKVREATKKGYAQAEIGDSINLERPSSQ 339

Query: 326 TRRGRVGKGIANTLLTGEEQGVVV--YDLYNRRKKDIVGTLTASGHNGNTTTGTFGISNG 383
```

```
            RRGRVGKGIANTL T  + GVVV  Y+  +++   + G L              G
Sbjct: 340 HRRGRVGKGIANTLTTSGQMGVVVASYEGEDKQVYQVAGVLID------------GQFYR 387

Query: 384 FRIRKLTPRECWRLQGFPDWAFDKASQVNSNSQLYKQAGNSVTVNVIAAIARRL 437
             RIR++TP+EC+RLQGFPDWAF+ A +V+SNSQLYKQAGNSVTV VIAAIA++L
Sbjct: 388 LRIRRITPKECFRLQGFPDWAFEAARKVSSNSQLYKQAGNSVTVPVIAAIAKKL 441
```

There is also homology to SEQ ID 2436:

```
Identities = 53/75 (70%), Positives = 62/75 (82%), Gaps = 1/75 (1%)

Query: 2   MKFLDLFAGIGGFRLGMEQAGHECIGFCEINKFARASYKVIHDTEGEIELHDITRVSDEF 61
            MKFLDLFAGIGGFRLG+    HECIGFCEI+KFAR SYK I++TEGEIE HDI +V+D+
Sbjct: 4   MKFLDLFAGIGGFRLGLINQCHECIGFCEIDKFARQSYKAIYETEGEIEFHDIRQVTDQD 63

Query: 62 IRGI-GSVDVICGGF 75
             R + G VD+ICGGF
Sbjct: 64 FRQLRGQVDIICGGF 78
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 797

A DNA sequence (GBSx0845) was identified in *S.agalactiae* <SEQ ID 2437> which encodes the amino acid sequence <SEQ ID 2438>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2585(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 798

A DNA sequence (GBSx0846) was identified in *S.agalactiae* <SEQ ID 2439> which encodes the amino acid sequence <SEQ ID 2440>. This protein is predicted to be arpR protein. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5070(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB09197 GB:U24159 orf12 [Bacteriophage HP1]
 Identities = 34/69 (49%), Positives = 47/69 (67%), Gaps = 1/69 (1%)

Query: 1   MTKTMTLEEKVEQWFIDRNLHE-ANPVKQFQKLIEETGELYSGIAKGKSEIIRDSLGDMQ 59
           M     L + +EQW  DRNL E + P KQF KL+EE GEL SG+AK K ++I+DS+GD
```

```
Sbjct: 1  MADLQQLIKNIEQWAEDRNLVEDSTPQKQFIKLMEEFGELCSGVAKNKPDVIKDSIGDCF 60

Query: 60 VVLIGIEQQ 68
           VV++ + +Q
Sbjct: 61 VVMVILAKQ 69
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 799

A DNA sequence (GBSx0847) was identified in *S.agalactiae* <SEQ ID 2441> which encodes the amino acid sequence <SEQ ID 2442>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.10   Transmembrane   13 -  29 (  10 -  36)

----- Final Results -----
              bacterial membrane --- Certainty=0.3039(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD21919 GB:AF085222 unknown [Streptococcus thermophilus
           bacteriophage DT1]
 Identities = 31/67 (46%), Positives = 49/67 (72%), Gaps = 1/67 (1%)

Query: 42  HQEADRVIIYVADNAGAEMFGKITDKEIIEGRHTVTAGAYGKFLVTEEQYNEITVGDDIP 101
           ++  + ++++ ADN   E+ GK+T K ++   +T+  GAYGKFLV++EQY+ + VGD+IP
Sbjct: 34  NRPVEAIVVHKADNF-VELHGKVTGKSMVGKLYTIDCGAYGKFLVSKEQYDSVQVGDEIP 92

Query: 102 DYLKGRG 108
            YLKGRG
Sbjct: 93  SYLKGRG 99
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 800

A DNA sequence (GBSx0848) was identified in *S.agalactiae* <SEQ ID 2443> which encodes the amino acid sequence <SEQ ID 2444>. This protein is predicted to be gene 17 protein. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5428(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA24397 GB:V01146 gene 1.7 [Bacteriophage T7]
 Identities = 30/72 (41%), Positives = 40/72 (54%)
```

```
Query: 47  DNVNYPSHYQGKYGLESIDVLRNFMTPEMLKGFYLGNALKYQLRYRKKNGLEDLKKARKN 106
           + V  PSHY     +E+I+V+   MT E  KG+  GN LKY+LR  KK+ L  L+K
Sbjct: 120 EGVTKPSHYMLFDDIEAIEVIARSMTVEQFKGYCFGNILKYRLRAGKKSELAYLEKDLAK 179

Query: 107 LDWLIEEMEKEK 118
            D+  E  EK K
Sbjct: 180 ADFYKELFEKHK 191
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 801

A DNA sequence (GBSx0849) was identified in *S.agalactiae* <SEQ ID 2445> which encodes the amino acid sequence <SEQ ID 2446>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1375(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 802

A DNA sequence (GBSx0850) was identified in *S.agalactiae* <SEQ ID 2447> which encodes the amino acid sequence <SEQ ID 2448>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0087(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10163> which encodes amino acid sequence <SEQ ID 10164> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF26608 GB:AF145054 ORF9 [Streptococcus thermophilus
           bacteriophage 7201]
 Identities = 99/148 (66%), Positives = 116/148 (77%), Gaps = 10/148 (6%)

Query: 5   MINNVVLIGRLTRDVELRYTPSNIANATFNLAVNRNFKNAAGDREADFINCVMWRQQAEN 64
           MINN VL+GRLT+D E +YT SNIA A+F+LAVNRNFK+A G+READFINCV+WRQQAEN
Sbjct: 1   MINNTVLVGRLTKDPEFKYTGSNIAVASFSLAVNRNFKDANGEREADFINCVIWRQQAEN 60

Query: 65  LANWTKKGMLIGITGRIQTRSYENQQGQRIYVTEVVADSFQILEKR----DNSTNQASMD 120
           LANW KKG LIGITGRIQTRSYENQQGQR+YVTEVVA++FQ+LE R       + N +
Sbjct: 61  LANWAKKGALIGITGRIQTRSYENQQGQRVYVTEVVAENFQMLESRAAREGGNANNSYSQ 120
```

```
Query: 121 DQLP------PSFGNSQPMDISDDDLPF 142
           Q+P        + N QP+DIS DDLPF
Sbjct: 121 QQVPNFARKNTEYSNKQPLDISSDDLPF 148
```

There is also homology to SEQ ID 1492.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 803

A DNA sequence (GBSx0851) was identified in *S.agalactiae* <SEQ ID 2449> which encodes the amino acid sequence <SEQ ID 2450>. This protein is predicted to be puff C4B protein. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1203(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10161> which encodes amino acid sequence <SEQ ID 10162> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 804

A DNA sequence (GBSx0852) was identified in *S.agalactiae* <SEQ ID 2451> which encodes the amino acid sequence <SEQ ID 2452>. This protein is predicted to be F5M15.19. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -2.34   Transmembrane    7 -  23 (   6 -  23)

----- Final Results -----
               bacterial membrane --- Certainty=0.1935(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 805

A DNA sequence (GBSx0853) was identified in *S.agalactiae* <SEQ ID 2453> which encodes the amino acid sequence <SEQ ID 2454>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4398(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10159> which encodes amino acid sequence <SEQ ID 10160> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 806

A DNA sequence (GBSx0855) was identified in *S.agalactiae* <SEQ ID 2455> which encodes the amino acid sequence <SEQ ID 2456>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2992(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 807

A DNA sequence (GBSx0856) was identified in *S.agalactiae* <SEQ ID 2457> which encodes the amino acid sequence <SEQ ID 2458>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4639(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB07758 GB:AP001520 unknown conserved protein [Bacillus halodurans]
 Identities = 65/184 (35%), Positives = 102/184 (55%), Gaps = 6/184 (3%)

Query: 1    MNIVEPLRDKDDIQAMKDYLSSWNEKYYMLFLLGINTGFRVGDILKLKVKDVQGWHIKVR 60
```

```
           M  V P RD D IQA+K  L   + + Y+LF +GINTG R+  +L LK+KDV
Sbjct: 1   MEYVVPFRDVDQIQAIKRSLKKKSPRDYLLFTIGINTGLRISQLLALKIKDVYDGQKPKD 60

Query: 61  EQKTGKYKSIKMTRPLKNELR---EFVKDKELHEYLFQSRVGKNKALSYKTVYWFLKRAA 117
             +    + + +   +K  L+    F++ +E H  LF S    ++ ++ +  Y  +K+AA
Sbjct: 61  YLQLESGEIVYLNDQVKKALQFYAHFIEFQEQH-CLFAS-TNPDQPMTRQHAYRIIKQAA 118

Query: 118 EDLGI-DNVGTHTMRKTFGYHYYKKYKNVADLMSLFNHSSPAVTLIYICVRQDELDTKMS 176
             +G+ D +GTHT+RKTFGYH Y++   ++ L   FNH +PA TL YI + ++E
Sbjct: 119 LQVGLTDQIGTHTLRKTFGYHAYRQGVALSLLQQRFNHQTPAQTLRYIDIAKNEQTIPRI 178

Query: 177 NFSL 180
           N +L
Sbjct: 179 NVNL 182
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 808

A DNA sequence (GBSx0857) was identified in *S.agalactiae* <SEQ ID 2459> which encodes the amino acid sequence <SEQ ID 2460>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3582(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 809

A DNA sequence (GBSx0858) was identified in *S.agalactiae* <SEQ ID 2461> which encodes the amino acid sequence <SEQ ID 2462>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2732(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 810

A DNA sequence (GBSx0859) was identified in *S.agalactiae* <SEQ ID 2463> which encodes the amino acid sequence <SEQ ID 2464>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1720(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 811

A DNA sequence (GBSx0860) was identified in *S.agalactiae* <SEQ ID 2465> which encodes the amino acid sequence <SEQ ID 2466>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2619(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10157> which encodes amino acid sequence <SEQ ID 10158> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 812

A DNA sequence (GBSx0861) was identified in *S.agalactiae* <SEQ ID 2467> which encodes the amino acid sequence <SEQ ID 2468>. This protein is predicted to be terminase large subunit. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2753(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC27181 GB:AF009630 putative terminase subunit [bacteriophage
           bIL170]
```

```
 Identities = 147/531 (27%), Positives = 261/531 (48%), Gaps = 26/531 (4%)

Query: 19  IRICKLTMKSIRRVERYKEQYLFKQEEADKRIEFIEEECSNTKGLAGKLRLALPQKVWLE 78
            I + K   K+I++  R  ++Y+++ +   + IE+IE+    T G   K++L   QK W E
Sbjct: 16  IELNKYMRKTIQKQIRIHKKYIYRYDRVTQAIEWIEDNFYLTTGNLMKIKLHPTQKYWYE 75

Query: 79  TTWGFYHTVEVTKTNPDTLEEYTDYEERRLIHEVPIIVPRGTGKTTLGSAIAEVGQIIDG 138
               G+           D ++E     +  LI+E+ + + RG+GK++L +       I+ G
Sbjct: 76  LMLGY-----------DMVDEKG--VQVNLINEIFLNLGRGSGKSSLMATRVLNWMILGG 122

Query: 139 EWGADIQLLAYSREQAGYLFNASRAMLSNEESLLHYMREADILRSTKQGILYETTNSLMS 198
            ++G +  ++AY   QA ++F+  R      ++L  Y  E  I +STKQG+ +    +
Sbjct: 123 QYGGESLVIAYDNTQARHVFDQVRNQTEASDTLRVY-NENKIFKSTKQGLEFTAFKTTFK 181

Query: 199 IKTSDYESLDGTNAHYNIFDEVHTYDDDFIKVVNDGSSRKRKNWITWYISTNGTKRDKLF 258
             +T+D     G N+  NIFDEVHTY +D  + VN GS +K+ NW + YI++ G KRD L+
Sbjct: 182 KQTNDTLRAQGGNSSLNIFDEVHTYGEDITESVNKGSRQKQDNWQSIYITSGGLKRDGLY 241

Query: 259 DKYYNIWVDILDDKIINDSVMPWIYQLDDVSEIHDPDMWQKAMPLLGITTEKETIARDIE 318
            DK    +    +++  ND     +Y L++  ++ D   W  A+PL+G   +   +  + E
Sbjct: 242 DKLVERFKS--EEEFYNDRSFGLLYMLENHEQVKDKKNWTMALPLIGDVPKWSGVIEEYE 299

Query: 319 MSKNDPAQQAELMAKTFNLPVNNYLAYFSNEECKGWSDKFDESLFVGDDERNARCVIGID 378
            +++ DPA Q + +A    LP+ +   YF+ ++ K    +F+ S+F        R  +GID
Sbjct: 300 LAQGDPALQNKFLAFNMGLPMQDTAYYFTPQDTK--LTEFNLSVF-----NKNRTYVGID 352

Query: 379 LSDVNDICSISFMVVRGEERHYLNKKFMPRHTIETLPKELRDKYTEWELSGMLHVHELDY 438
            LS + D+ ++SF+     + +     F  R   E L  E ++ +TE+   G L + + +Y
Sbjct: 353 LSLIGDLTAVSFVCELEGKTYSHTLTFSVRSQYEQLDTEQQELWTEFVDRGELILLDTEY 412

Query: 439 NDQAYIFEELRQFMSDNRILPVAVGYDRYNARELIRLFNDYYGDICHDIPQTVK---SLS 495
             +   +   +  F S        +GYD      L  L   Y+ D   D  + ++   S++
Sbjct: 413 INVNDLIPYINDFRSKTGCRLRKIGYDPARYEILKGLIERYFFDKDGDNQRAIRQGFSMN 472

Query: 496 NPLKVYKEKAKMGKIIFDDPVATWNHANVRVKIDANNNIFPNKEKAKEKID 546
            + +K+ K K    K+I +  V  W   N  VKI  + +    K+  K+KID
Sbjct: 473 DYIKLLKSKLVENKLIHNQKVMQWALNNTAVKIGQSGDYMYTKKLEKDKID 523
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 813

A DNA sequence (GBSx0862) was identified in *S.agalactiae* <SEQ ID 2469> which encodes the amino acid sequence <SEQ ID 2470>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3319(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB41469 GB:L35061 orfL4 [Bacteriophage phi-41]
 Identities = 86/374 (22%), Positives = 166/374 (43%), Gaps = 38/374 (10%)

Query: 12  FARIFRPNNRKSTRTYLQRSISYWRRNSIYLDNIYNKISTDTAQLRFKHVKITRNPGGVD 71
            F+R    N+ +    +   ++ Y    S ++ NI+NKI+ +  ++ F HVK ++   G D
Sbjct: 10  FSRGKLNNDTQRVTAWQNEAVEY---TSAFVTNIHNKIANEITKVEFNHVKYKKSDVGSD 66

Query: 72  SMVWYEHSDLAEVLTVSPNPLEVPVVFWSNVTRAMLRDGVAVVVPRW--KNGRLVEIWLA 129
```

```
           +++    SDL EVL  S       + FW  V + +L      + P +  K G LV++  A
Sbjct: 67  TLISMAGSDLDEVLNWSSKGERNSMEFWQKVIKKLLTTRYIDLYPIFDRKTGDLVDLLFA 126

Query: 130 KKTVTWTAESVELMLDDVAVELPLTDVWVFENPKLNVTAQLNQITELIDINLNALTEKLS 189
                +  E +  ++    +                     N+ T ++D  L  +  KL
Sbjct: 127 DNKKEYKPEELVRLISPFYI---------------------NEDTSILDNALAGIQTKLE 165

Query: 190 DGNSSLRGFLKLPT---KAADEHLKQQARDRVDSMLDLAKNGGIAYLEQGEEFQELSKDY 246
            G   ++G LK+        D+  K +A   + +M +++   G+   +   E  EL KDY
Sbjct: 166 QGK--MKGLLKINAFIDTDNDQEFKDKAMLTIKNMQEMSNYNGLTPTDNKTEIVELKKDY 223

Query: 247 STASKEELEFLKSQLYNAHGINEKLFTCDYTEEQYRAYYSSVMKLYQRVYSEEINRKYFT 306
           S  +K+E++ +KS+L   + +NE +     ++EQ   +Y+S +        +E+  K  +
Sbjct: 224 SVLNKDEIDLIKSELLTGYFMNENILLGTASQEQQIYFYNSTIIPLLIQLEKELTYKLIS 283

Query: 307 KTAR--TQGN----KLLVFFDMADMISFKDLVEGGFKSKYAGLMNSNEFRETYLGLPGYE 360
              R   +GN    +++V   +    + K+L++   ++    +   N+     +G    E
Sbjct: 284 TNRRRVVKGNLYYERIIVDNQLFKFATLKELIDLYHENINGPIFTQNQLL-VKMGEQPIE 342

Query: 361 GGEVFETNLNAVRI 374
           GG+V+  NLNAV +
Sbjct: 343 GGDVYIANLNAVAV 356
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 814

A DNA sequence (GBSx0863) was identified in *S.agalactiae* <SEQ ID 2471> which encodes the amino acid sequence <SEQ ID 2472>. This protein is predicted to be a prohead protease. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3496(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF31089 GB:AF069529 protease [Bacteriophage HK97]
 Identities = 52/142 (36%), Positives = 73/142 (50%), Gaps = 11/142 (7%)

Query: 21  FEAYASTYDNTDREGDVMAKGCFDNTLKSKA-VVPMCLNHDR-NCVIGKHE-LSVDEKGL 77
            FE YAS ++NTD +GD++  G F N L ++   V M  NH      +GK + L+ DEKGL
Sbjct: 26  FEGYASVFNNTDSDGDIILPGAFKNALANQTRKVAMFFNHKTWELPVGKWDSLAEDEKGL 85

Query: 78  RTRSTFNLSDPEAKKTYDLMKMGALDSLSIGFFI--KDYEPIDAKQPYGGWIFKEVE-IF 134
              R         A      M+ G ++ +S+GF +   DY  I       G IFK ++ +
Sbjct: 86  YVRGQLTPGHSGAADLKAAMQHGTVEGMSVGFSVAKDDYTIIPT-----GRIFKNIQALR 140

Query: 135 EISVVTVPANPQATVDNIKEFD 156
            EISV T PAN QA +  +K  D
Sbjct: 141 EISVCTFPANEQAGIAAMKSVD 162
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 815

A DNA sequence (GBSx0864) was identified in *S.agalactiae* <SEQ ID 2473> which encodes the amino acid sequence <SEQ ID 2474>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2247(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10155> which encodes amino acid sequence <SEQ ID 10156> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC27185 GB:AF009630 16 [bacteriophage bIL170]
 Identities = 70/249 (28%), Positives = 121/249 (48%), Gaps = 23/249 (9%)

Query: 51  LEQLKTDAESLVSQATA--IKETIAGLDSDIEETEEELSK-AAKIIK---------EKQK 98
            L +LK +  SL SQ     +K  I  L   ++E E+ LS+ + +IIK         EK K
Sbjct: 13  LAELKENNVSLKSQINGFEVKNAIEDLPK-VQELEKTLSENSIEIIKIENELNAQEEKPK 71

Query: 99  GNTPM-DYLKTKAAALDFVRILMDNEGSANSARKAWEANLVEKGV--TNLTKILPEPVLI 155
            G   M ++++++ A  +F  +L  N G +    + AW A L E GV  T+ T  LP   ++
Sbjct: 72  GKAKMTNFIESQNAVTEFFDVLKKNSGKSE-IKNAWNAKLAENGVTITDTTFQLPRKLVE 130

Query: 156 AIQDAFTNYNGILN--HVSKDPRYAVRVALQTQVSQAKGHKAGKTKKDEDFTFLDFTINS 213
            +I  A  N N +    HV+      V  +  +  ++A+ HK G+TK ++  T    T+
Sbjct: 131 SINTALLNTNPVFKVFHVTNVGALLVSRSFDSS-AEAQVHKDGQTKTEQAATLTIDTLEP 189

Query: 214 ATVY-IKYAFEYSDLKKDTTGAYFNYVMKELAQGFI-RTIERAVVIGDGKSN-SAEDKIT 270
              VY ++   E     + +    +N ++ EL Q  + + ++ A+V GDG +   + DK
Sbjct: 190 VMVYKLQSLAERVKRLQMSYSELYNLIVAELTQAIVNKIVDLALVEGDGSNGFKSIDKEA 249

Query: 271 EIKSIAEET 279
            ++K I + T
Sbjct: 250 DVKKIKKIT 258
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 816

A DNA sequence (GBSx0865) was identified in *S.agalactiae* <SEQ ID 2475> which encodes the amino acid sequence <SEQ ID 2476>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3068(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 817

A DNA sequence (GBSx0866) was identified in *S.agalactiae* <SEQ ID 2477> which encodes the amino acid sequence <SEQ ID 2478>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0437(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 818

A DNA sequence (GBSx0867) was identified in *S.agalactiae* <SEQ ID 2479> which encodes the amino acid sequence <SEQ ID 2480>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3181(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10153> which encodes amino acid sequence <SEQ ID 10154> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 819

A DNA sequence (GBSx0869) was identified in *S.agalactiae* <SEQ ID 2481> which encodes the amino acid sequence <SEQ ID 2482>. This protein is predicted to be a major structural protein. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3364(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA74331 GB:L33769 unidentified ORF28; putative [Bacteriophage
          bIL67]
 Identities = 55/201 (27%), Positives = 84/201 (41%), Gaps = 18/201 (8%)

Query: 9   EVTHGNANGF-YAKIAKTDAGALDLQKPYPFTGLRSTSFETSQESNAYYAD-NVEHVRLQ 66
           E+THG   G  +  +  +  G        P  GLR    ++ QE+  +YA  N  +  +
Sbjct: 8   ELTHGLGYGVVFTDLTGSKTGI-------PIAGLRGIETDSKQENKNFYAGFNAPYRTIA 60

Query: 67  GKKSTEGSITTYQIPKQFMIDHLGKKLTNSTPPALIDTGVNTN-FIWGYAETVTDEFGAE 125
           G K T+  + +Y +P  F    LG     S    L D   N   + + YAE   D+ G
Sbjct: 61  GAKDTQIKVKSYDLPDDFATHALG---FGSVQGFLTDDVANYKPYGFAYAERYRDDDGTG 117

Query: 126 IEEFHIWTNVKASAPKGSTSTDETSATPKEIEIPCTASPNNFIVDSEKKPVSEIVWRDDS 185
            +    + +V+A+ P  +   DE S T KE E   T +  +F +  +K+   +    D
Sbjct: 118 YKA-TFYPSVQATTPSDTAEADEESPTGKEYEHEATVTTGDFTLGDKKRLFVKFKVSDTE 176

Query: 186 KGT-VRGK---FDKLFADKSP 202
             T   GK   F KLF D  P
Sbjct: 177 LATGTSGKALAFKKLFTDLKP 197
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 820

A DNA sequence (GBSx0870) was identified in *S.agalactiae* <SEQ ID 2483> which encodes the amino acid sequence <SEQ ID 2484>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2531(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 821

A DNA sequence (GBSx0871) was identified in *S.agalactiae* <SEQ ID 2485> which encodes the amino acid sequence <SEQ ID 2486>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2972(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 822

A DNA sequence (GBSx0872) was identified in *S.agalactiae* <SEQ ID 2487> which encodes the amino acid sequence <SEQ ID 2488>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3860(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 823

A DNA sequence (GBSx0873) was identified in *S.agalactiae* <SEQ ID 2489> which encodes the amino acid sequence <SEQ ID 2490>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.22   Transmembrane  605 - 621 ( 569 - 631)
   INTEGRAL    Likelihood = -8.12   Transmembrane  583 - 599 ( 569 - 604)

----- Final Results -----
               bacterial membrane --- Certainty=0.6689(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAB70053 GB:AF011378 unknown [Bacteriophage sk1]
 Identities = 159/709 (22%), Positives = 285/709 (39%), Gaps = 112/709 (15%)

Query: 128 SILNLNKELDNVAKELDIVNQKLELDPDNVELAEQKMKLLGKQSELAGDKVQELKKKQAA 187
           S+  +N  +  +  E   +   L+LDP N +   Q  K L  Q  L+ DK  +LK++ ++
Sbjct: 21  SLKGVNTAMSGLRGEAKNLRDALKLDPTNTDKMAQLQKNLQTQLGLSRDKATKLKQELSS 80

Query: 188 LGDEK-IGTEEWRQLQNEIGQAEVEVLKIDRAMDILGESSRSATGDI--KEATSYLRADV 244
           +      G ++W QL  ++G AE +  +++  +  +  +  S + DI  K  T  + + +
Sbjct: 81  VDKSSPAGQKKWLQLTRDLGTAETQANRLEGEIKQVEGAISSGSWDIDAKMDTKGVNSGI 140

Query: 245 MMDVADKAG----------QIGQKMVDAGKMTVDAWSEIDEALDTVTTKTGLTGD----- 289
               +  +G          QIG   V A    +  W  + +A+DT      L
Sbjct: 141 DGMKSRFSGLREIAVGVFRQIGSSAVSAVGNGLKGW--VSDAMDTQKAMISLQNTLKFKG 198

Query: 290 -------ALAELQEIAKDIATG------MPTSFQNAGD----AVGEL------NTQFGLT 326
                      +Q +AKD          + T+F   GD    AVG+       N  FG T
Sbjct: 199 NGQDFDYVSKSMQTLAKDTNANTEDTLKLSTTFIGLGDSAKTAVGKTEALVKANQAFGGT 258

Query: 327 GEKLKSASELL--------IKYAEINE-TD--------ISSSAISAKQAIEAYG--LTAE 367
           GE+LK   +          +    IN+ TD        + S+ +    A++ YG   +A
Sbjct: 259 GEQLKGVVQAYGQMSASGKVSAENINQLTDNNTALGSALKSTVMEMNPALKQYGSFASAS 318

Query: 368 DLGMV----LDNVTKAAQDTGQSVDTIVQKAIDGAPQIKGLGLSFEEGA------ALIGK 417
```

```
           + G +    LD    +     G   T +  A D   +   L L     A      ++I K
Sbjct: 319 EKGAISVEMLDKAMQKLGGAGGGAVTTIGDAWDSFNETLSLALLPTLDALTPIISSIIDK 378

Query: 418 FEKSGVDSSAALSSLSKAAVIYAKD--GKTLTDGLNETVSAIQNSTSET--EALSIASEI 473
               G  +  AL S+ K    Y K+  G    +G   ++S I +    T    LSI  ++
Sbjct: 379 MAGWGESAGKALDSIVK----YVKELWGALEKNGALSSLSKIWDGLKSTFGSVLSIIGQL 434

Query: 474 FGSKAAPRMVDAIQRGAFSFDDLAEAAKSSSGTVSTTFDETLDPIDKLTQYSNQAKEGMA 533
             S A    +D+         + A + ++ S T++       D I K+  ++ +  E
Sbjct: 435 IESFAG---IDS------KTGESAGSVENVSKTIANLAKGLADVIKKIADFAKKFSESKG 485

Query: 534 ELGGKLLETVIPALEPLMGMLESSVNWFTSLNETDQ-QTIVILGLVTTAVMMLLGAIAPL 592
            +    L+T + AL       +      T+++   + QT + G         + AI P
Sbjct: 486 AID--TLKTSLVALTAGFVAFKIGSGIITAISAFKKLQTAIQAGTGVMGAFNAVMAINPF 543

Query: 593 VIAIGAIGAPVGIVVAAIV-GAIAVITLIIQAIMNWGAITEWLQSTWDSCAA-------W 644
           V         +GI +AAIV G +   T        W +  ++L+S WD   +       W
Sbjct: 544 VA--------LGIAIAAIVAGLVYFFTQTETGKKAWASFVDFLKSAWDGIVSFFSGIGQW 595

Query: 645 LSELWTNIVTTATTAWSNFTAWLSGLWSSVVSTGQSLWSSFTSSLSNIFSSLITGAQSLW 704
            +++W   V  A   W     W SG+   V    Q++W+  T+  + ++++++TG Q+ W
Sbjct: 596 FADIWNGAVDGAKGIWQGLVDWFSGIVQGV----QNIWNGITTFFTTLWTTVVTGIQTAW 651

Query: 705 SSFTSTLSNLWSGLVSTGSNLFNNLSSTISGIFNGILSTASNIWNSIKS 753
           +  T   + LW G+V+  + +F  +SS ++G +N  ++T   + +  KS
Sbjct: 652 AGVTGFFTGLWDGIVNVVTTVFTTISSLVTGAYNWFVTTFQPLISFYKS 700
```

There is also homology to SEQ ID 2492.

A related GBS gene <SEQ ID 8663> and protein <SEQ ID 8664> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:    -13.98
GvH: Signal Score (-7.5): -2.78
     Possible site: 16
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value: -14.22 threshold:  0.0
   INTEGRAL    Likelihood =-14.22   Transmembrane  605 - 621 ( 569 - 631)
   INTEGRAL    Likelihood = -8.12   Transmembrane  583 - 599 ( 569 - 604)
   PERIPHERAL  Likelihood =  4.45     539
 modified ALOM score:   3.34

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6689(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear)
```

The protein has homology with the following sequences in the databases:

```
27.1/51.7% over 981aa
                         Bacteriophage sk1
  GP|2392838| unknown Insert characterized

ORF00471(328 - 2976 of 3333)
GP|2392838|gb|AAB70053.1||AF011378(9 - 990 of 999) unknown {Bacteriophage sk1}
%Match = 7.3
%Identity = 27.1  %Similarity = 51.7
Matches = 164  Mismatches = 275  Conservative Sub.s = 149

243       273       303       333       363       393       423       453
MSINQEEKKTLSNADLLSVMSD*KERRKSMTETFEGLYVKFGANTVEFDRSVKGINTALSSLKKDFNNINRQLKMDPDNV
                                :   : |:   || :|: |:||:|||:| |: :  |:   ||:|| |
                              MASNATFEVEIYGNTTKFENSLKGVNTAMSGLRGEAKNLRDALKLDPTNT
                                      10        20        30        40        50
```

```
483       513       543       570       600       630       660       690
DLLNRKLVNLQEQARVGAIKIAELKKQQKALGESE-VGSAQWNKLQLEIAKVESQMKIVDKAMESTKKHIEDVGDPKSIL
| : :   ||| |  :   |  :||::  :: :|   |  :| :|  ::   |:|   ::  :: :
DKMAQLQKNLQTQLGLSRDKATKLKQELSSVDKSSPAGQKKWLQLTRDLGTAETQANRLEGEIKQVE-------------
      60        70        80        90       100       110

1053      1083      1113      1143      1167      1197      1227
NLNKELD~~~~DVMMDVADKAGQIGQKMVDAGKMTVDAWSEIDEALDTVTTKTGLTG--DALAELQEIAKDIATGMPTSF
                               | ::  :| :||  :||    :|: |    :: |:|||  :   : :|
-------~~~~--------------------GAISSGSW-DIDAKMDTKGVNSGIDGMKSRFSGLREIAVGVFRQIGSSA
                               120      130       140       150       160

cag                                                                            g
aac                                                                            g
gtt                                                                            t
                                                                            1239
QNA----------------------------------------------------------------------------G
 :|
VSAVGNGLKGWVSDAMDTQKAMISLQNTLKFKGNGQDFDYVSKSMQTLAKDTNANTEDTLKLSTTFIGLGDSAKTAVGKT
          180       190       200       210       220       230       240

1269      1299      1329      1359      1389      1416      1446      1476
DAVGELNTQFGLTGEKLKSASELLIKYAEINETDISSSAISAKQAIEAYG-LTAEDLGMVLDNVTKAAQDTGQSVDTIVQ
:|: : |  || |||:||                     : |   : | | ::||::  : || |
EALVKANQAFGGTGEQLKGV-------------------VQAYGQMSASGKVSAENINQLTDNNT---------------
          260                          270       280       290

1506      1536      1566      1596      1626      1656      1686      1716
KAIDGAPQIKGLGLSFEEGAALIGKFEKSGVDSSAALSSLSKAAVIYAKDGKTLTDGLNETVSAIQNSTSETEALSIASE
                                                     |   |  ||  :  :  :   :::|||
----------------------------------------------------ALGSALKSTVMEMNPALKQYGSFASASE
                                                          300       310

1746                1794      1824      1854      1884      1914      1944
IFGSKAAPRMVDAIQR----GAFSFDDLAEAAKSSSGTVSTTFDETLDPIDKLTQYSNQAKEGMAELGGKLLETVIPALE
  |: :   :  |:|:    |  :   : :|  | : |:|  :  ||| :  :         |  |  || |::::  ::
-KGAISVEMLDKAMQKLGGAGGGAVTTIGDAWDSFNETLSLALLPTLDALTPIISSIIDKMAGWGESAGKALDSIVKYVK
       330       340       350       360       370       380       390

1974      2004      2034      2064
PLMGMLESSVNWFTSLNETDQQTIVILGLVTTAVMMLLGAIAPL--------------------~~~~------------
 | | || :   ::||::        :| | : :  |: : | :
ELWGALEKN-GALSSLSKIWDGLKSTFGSVLSIIGQLIESFAGIDSKTGESAGSVENVSKTIAN~~~~FKKLQTAIQAGT
        410       420       430       440       450       460

2082      2112      2139      2169      2199                2238      2268
--------VIAIGAIGAPVGIVVAAIV-GAIAVITLIIQAIMNWGAITEWLQSTWD-------SCAAWXSELWTNIVTTA
        |:||      :|| :|||| | :   |        | :  ::|:| ||           | :::|   |  |
GVMGAFNAVMAINPF-VALGIAIAAIVAGLVYFFTQTETGKKAWASFVDFLKSAWDGIVSFFSGIGQWFADIWNGAVDGA
       540        550       560       570       580       590       600

2298      2328      2358      2388      2418      2448      2478
TTAWSNFTAWLSGLWSSVVSTGQSLWSSFTSSLSNIFSSLITGAQSLWSSFTSTLSNLWSGLVSTGSNLFNNLSS-----
   |  :  |:||:   |    |::|:  |: :: ::::::|| |: |:  |  :: || |:|:  : :|  :||
KGIWQGLVDWFSGIVQGV----QNIWNGITTFFTTLWTTVVTGIQTAWAGVTGFFTGLWDGIVNVVTTVFTTISSLVTGA
        620           630       640       650       660       670       680

2496      2526      2556      2586      2616
---------------~~~~--------------------TISGIFNGILSTASNIWNSIKSTISNAIDGAKNAVSNGVNA
                                        |:|:|||| |  |:|:  :|:|| : : |  : ::|   :
YNWFVTTFQPLISFY~~~~KNIVSGVFEAFGNFASNAWNAITGVFNGIGSFFSDIFGGVKNTIDSVLGGVTDTINNIKGS
                    870       880       890       900       910       920

2646      2676      2706      2736      2766      2796      2826      2856
IKNLFNFQIKWPHIPLPHFRVSGSANPLDWLKGGLPSIGIDWYAKGGIMTKPTLFGMNGNRAMVGGEAGAEAILPLNKST
|
                                                         :   : |||      :: |
I--------------------------------------------------------DWVASKVGGLFKGSMVVGLTDVN
```

```
                                                         930       940       950

2886      2916      2946      2976      3006      3036      3066      3096
LGAIGQSIANTMNTSNNINVNFSGVTIREEADLNRLANVVGNRIAEELQRKTNLRGGMA*QKSMNLPLTV*KHHLLSVMY
| : |  ::     :|:|    |   |     |:: ||  :
LSSSGYGLSTNSVSSDNRTYNTFNVQGGAGQDVSNLARAIRREFELGRA
        960       970       980       990
```

SEQ ID 8664 (GBS58) was expressed in and purified from *E.coli* as a GST fusion. The purified protein is shown in lane 10 of Figure 193.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 824

A DNA sequence (GBSx0874) was identified in *S.agalactiae* <SEQ ID 2493> which encodes the amino acid sequence <SEQ ID 2494>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2732(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 825

A DNA sequence (GBSx0875) was identified in *S.agalactiae* <SEQ ID 2495> which encodes the amino acid sequence <SEQ ID 2496>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2467(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10151> which encodes amino acid sequence <SEQ ID 10152> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10935> which encodes amino acid sequence <SEQ ID 10936> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2497> which encodes the amino acid sequence <SEQ ID 2498>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
                bacterial cytoplasm --- Certainty=0.2136(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 55/240 (22%), Positives = 92/240 (37%), Gaps = 20/240 (8%)

Query: 4   INELTIDGVKTSSFKCDVLVETRPNVIVSSS--KTALLEHDGISGAVVQSNRHRGLIEKP 61
           I ++ ID   TSS    VL       I+S S  +      +G S  +   N  +  I
Sbjct: 2   IPKVIIDDFDTSSIPNCVLTGYDVGDILSPSFVENEAYGMNGTSRELESYNESKPTIM-- 59

Query: 62  YHITLIEPSDEEIYRFSALLNREKFW-LENEQEPTIRLWCYKVDSFEIGKDEFGAWVVDV 120
           +H++  + +   I     L  + +FW + N       ++ Y   S +I      +W V +
Sbjct: 60  WHLSTFDDAVNLINHLDGLSKKIEFWHIPNS------IYYYDCLSVKINAVTMSSWRVTL 113

Query: 121 TFICHPTKFFKTTDIQTLTGNGVLRVQGSALAFPKITVVGQSASETSFTIGNQVIKLEKL 180
               +P ++ K      + GNG +   G+  + PKI V G    + + TIG QV++L  L
Sbjct: 114 KLALYPFRYAKGVSDVVIAGNGNINNAGNVFSEPKIVVEG--TGKGTLTIGKQVMEL-NL 170

Query: 181 SESLVMTNDPDNPSFKTASGKL---IKWAGDFITVDTAKGQNVGVVLGAGITSLKFETVW 237
           S    +           A G +   I+  G F  +        G+ +  GIT       W
Sbjct: 171 SGKATIECKHGQQCVYDAEGNVKNSIRIRGSFFEIQPG---TQGIAVSGGITRTIISPRW 227
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 826

A DNA sequence (GBSx0876) was identified in *S.agalactiae* <SEQ ID 2499> which encodes the amino acid sequence <SEQ ID 2500>. This protein is predicted to be PblB. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.00   Transmembrane  952 - 968 ( 952 - 968)

----- Final Results -----
               bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG18640 GB:AY007505 PblB [Streptococcus mitis]
 Identities = 145/542 (26%), Positives = 255/542 (46%), Gaps = 52/542 (9%)

Query: 1   MLFLLDANVRTVKWNGIPLHEASSAIVKEETNGDFYLTVRYPITDSGIYQLIKEDMLIKS 60
           M++L + N         PL+ A +  + +E N  + LT R+P +D  +++ +KE+  +K+
Sbjct: 1   MIYLTNGNT--------PLNAAYADKISQEANSTYQLTFRFPTSDV-LWEKLKEETFLKA 51

Query: 61  PVPVLGAQLFRIKKPIENDDSMDITAYHVSDDIMKRSITPVSVVGQGCAMALSQMVQNAK 120
              + G Q F I +  +    + + A  V   +    I P+S+     + ALS+   +
Sbjct: 52  D-DLHGEQDFVIFEVQKKHGYIQVYANQVMTLLNNYVINPISLDRATGSTALSRFAGSI- 109

Query: 121 TGLGDFSFTSDIMDSRTFNTTETETLYSVLMDGKHSIVGTWEGELVRDNFALSIKRSRGA 180
           T    FSF SDI +  TFNT     + +   D KHSI+G W G+LVR  + + + ++ G+
Sbjct: 110 TRYNTFSFFSDIDERHTFNTDSVNAMVAFTKD-KHSILGQWGGDLVRHGYQVRLLKNGGS 168

Query: 181 DRGVVITTHKNLKSYQRTKNSQGVVTRIHARSTFKPDGAE-DEVTLRVSVDSPLINSYPY 239
           +   +    KNL SYQ   +++ + TRI  ++T K +G +  +    V VDSPL+N Y
Sbjct: 169 ENESLFMYKKNLSSYQHKTSTKSLKTRITFKATVKGEGEKAPDRKFSVVVDSPLVNKYSQ 228

Query: 240 INEKEYENNNAETVED--LRKWAEAKFTNEGIDKVSDAIEIEAYELDGQVVNLGDTVNLK 297
```

```
           I E   E N+ +  ++  LRK+ E  F     D + D++EI+        V + D V+L
Sbjct: 229 IYEDVIEVNDQDVKDEVGLRKYGEQYFRTTLCDMLEDSLEIQVEGKSDVPVQIFDIVSLF 288

Query: 298 SRKHSADLYKKAIAYEFNALTEEYISITFDDKPGVGGSGVSSGLSN-VADAILVASATAQ 356
             +   D+ KK   Y ++ + ++ +SI F    G   SG+S+ LSN V+DA+   +   Q
Sbjct: 289 HDRFKMDVRKKITKYTYSPMAKKLLSIGF----GQFKSGLSNMLSNAVSDAVKNETQHLQ 344

Query: 357 D---VAVQRAVKNANAAFDAEFGKTKTKINDDIEIAKAKVESFKSELSNRMDNQLLP--- 410
                 + + +KNA+ AFD +  +   +  D +  AKAK E  K  L+  +D +
Sbjct: 345 GQFATQLGKEIKNADLAFDRKKEELVNQFTDGLNAAKAKAEEVKKSLTETIDQRFRDFDS 404

Query: 411 ----------------------LATEAKNLASQAQADLTRKEIELRAELNRQVTSTEAVK 448
                                 LA EAK ++ QA+  +  K  E +  ++ + TS  +
Sbjct: 405 TGLNEIKQKAEEALQRVGANTLLAQEAKQISEQARQQMDSKFAEYKQSVDGRFTSLSSQL 464

Query: 449 ISLTNLSHNMDIIKQKALNDLRDAETRLKEADSVQQLATKRVEDKLTGLSTKLESFSVGG 508
               NL   +D  + +  ++L +      E+D  +++A   + ++L  +     S +VGG
Sbjct: 465 AGKANL---IDFQRVQEKSNLYERIIGSSESDIAEKVARMTLTNQLFQVEVGKYS-AVGG 520

Query: 509 YN 510
            N
Sbjct: 521 PN 522
 Identities = 47/183 (25%), Positives = 83/183 (44%), Gaps = 22/183 (12%)

Query: 867  VTTLRVTKGTIPADWSPSPDDLKAYSDTKLEQTANEIKASVTSLDHKTLKQTDITMTSEG 926
            +T L   +GT    W P+P+D    +D  LE T                 QT +T+
Sbjct: 667  MTELDFYEGTTDRRWQPAPEDATLETDKTLEAT-----------------QTKLTLLQGS 709

Query: 927  IVLRAGKTSNDVARAIGSYFKVTPDAIALFSSLIKVSGNMLVDGSVTSRKLVTGAVETGH 986
              ++   TS   A +I S    T + I + +  I++ G  L+D  +T+       +  G
Sbjct: 710  FAIQ-NLTS---AGSIVSQINATNNQILIEAEKIRLKGKTLLD-ELTAIDGYFKRLFVGE 764

Query: 987  VKAGAITGVLLAAEAVTAEKLKVDQAFFNKLMANDAYLKQLFAKSAFITQVQSVTISASQ 1046
                 +   ++ ++ +TA+KL +DQA     +++D +   L AK AFI +++SV +SA+
Sbjct: 765  GTFAKLNAEIIGSKTITADKLIMDQAMARLFVSSDIFTDTLAAKEAFINKLRSVVVSATL 824

Query: 1047 ISG 1049
              G
Sbjct: 825  FEG 827
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2501> which encodes the amino acid sequence <SEQ ID 2502>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2445(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 143/552 (25%), Positives = 251/552 (44%), Gaps = 43/552 (7%)

Query: 11  TVKWNGIPLHEASSAIVKEETNGDFYLTVRYPITDSGIYQLIKEDMLIKSPVPVLGAQLF 70
           ++K +  PL  A    + +E N D+ L  +YP        LIK+ +++++   + G+QLF
Sbjct: 3   SIKDDNTPLVAAFEDEITQEANSDYKLNFKYPAKHE-YRPLIKKGIILEAD-DLHGSQLF 60

Query: 71  RIKKPIENDDSMDITAYHVSDDIMKRSITPVSVVGQGCAMALSQMVQNAKTGLGDFSFTS 130
           RI +  +    +++ A  V+DD+   +I  +SV        +S++  + K     FSF S
Sbjct: 61  RIFEITKRHGYINVYANQVADDLNGYAIDTISVDRVQGMTVMSELAGSIKRE-HPFSFFS 119

Query: 131 DIMDSRTFNTTETETLYSVLMDGKHSIVGTWEGELVRDNFALSIKRSRGADRGVVITTHK 190
           DI    TFN ++   +   L +GKHSI+G W GELVR+ + +++ +  G D   +    K
Sbjct: 120 DIDGRHTFNQSDVSVM-DALANGKHSIMGQWGGELVRNKYQINLLKKAGKDTETLFMYKK 178
```

```
Query: 191 NLKSYQRTKNSQGVVTRIH----------ARSTFKPDG------AEDEVTLRVSVDSPLI 234
           NLKSY+ T   +G+V+ +H            +    DG      +  + T+RVSV+S L
Sbjct: 179 NLKSYEETDTIKGLVSILHLVAEVEEEHEVETREASDGNIGHSESPKKKTIRVSVESKLK 238

Query: 235 NSYPYINEK--EYENNNAETVEDLRKWAEAKFTNEGIDKVSDAIEIEAYELDGQVVNLGD 292
           +++P I EK  + ++ + +T EDL  + +  F     D   ++++I+        V L D
Sbjct: 239 DTHPIIVEKTIKVQDQDVKTEEDLLAYGKKYFEKTLCDIPGNSLKIDVTNNYEGAVRLFD 298

Query: 293 TVNLKSRKHSADLYKKAIAYEFNALTEEYISITFDDKPGVGGSGVSSGLSNVADAILVAS 352
           T  +    +  DL  +   Y F  +     SI F    G   + ++  +SN  D  + S
Sbjct: 299 TAIVFHELYDRDLRMQITGYRFAPMANRLKSIIF----GEIKTNLAKQISNQIDNKVAES 354

Query: 353 ATAQDVA----VQRAVKNANAAFDAEFGKTKTKINDDIEIAKAKVESFKSELSNR-MDNQ 407
               D A    +Q+ + NAN  FD +  K + +I D I+ A+A  E   +E++ + ++ +
Sbjct: 355 TAQHDAAFEAKLQKQIDNANRIFDTKEAKLREEIEDGIKKAEANAEVKVAEVNAKVLEAE 414

Query: 408 LLPLATEAK-----NLASQAQADLTRKEIELRAELNRQVTSTEAVKISLTNLSHNMDIIK 462
            L  A + +     + A   + D  +K  E R  L       + +  +L     + D +
Sbjct: 415 ELAKAVDERLKKFLSDADTKEQDFDKKLEEFRTSLKDLEVDEKQIDDALAKAGFSKDSLA 474

Query: 463 QKALNDLRDAETRLKEADSVQQL-ATKRVEDKLTGLSTKLESFSVGGYNYVIDGGEPKEL 521
                    +ET    A+ V     T    ++L G + K+ +F   GY  +   GE  E
Sbjct: 475 DIKAKLEDTSETATVTANIVGSTGGTFYNRNRLDGDTDKVITFE-QGYIDIAHNGEGFE- 532

Query: 522 MANFYGKTYDIN 533
                GKTY I+
Sbjct: 533 ----EGKTYTIS 540
```

A related GBS gene <SEQ ID 8665> and protein <SEQ ID 8666> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 3
SRCFLG: 0
McG: Length  of  UR:   11
     Peak Value of UR:   1.54
     Net Charge of CR: 1
McG: Discrim Score:     -3.43
GvH: Signal Score (-7.5): -5.44
     Possible site: 58
>>> Seems to have no N-terminal signal sequence
Amino Acid Composition: calculated from 1
ALOM program   count: 1 value:  -0.00 threshold:  0.0
   INTEGRAL    Likelihood = -0.00   Transmembrane  897 - 913 ( 897 - 913)
   PERIPHERAL  Likelihood =  1.48     932
 modified ALOM score:   0.50
icml HYPID: 7  CFP: 0.100

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
32.8/53.9% over 503aa
EGAD|33685| hypothetical protein Insert characterized
  EGAD|71773|76294 hypothetical protein { } Insert characterized
  SP|P15317|YHYA_BPH44 HYPOTHETICAL 65 KDA PROTEIN IN HYALURONIDASE REGION. Insert
characterized
  GP|215054|gb|AAA98102.1||M19348 ORF {Streptococcus pyogenes phage H4489A} Insert
characterized
  PIR|B30566|B30566 hypothetical protein - phage H4489A Insert characterized

ORF00870(1957 - 3777 of 4272)
```

```
EGAD|33685|35003(37 - 540 of 593) hypothetical protein {Streptococcus pyogenes}
EGAD|71773|76294 hypothetical protein { } SP|P15317|YHYA_BPH44 HYPOTHETICAL 65 KDA PROTEIN
IN HYALURONIDASE REGION. GP|215054|gb|AAA98102.1||M19348 ORF {Streptococcus pyogenes phage
H4489A} PIR|B30566|B30566 hypothetical protein - Streptococcus pyogenes phage H4489A
%Match = 4.4
%Identity = 32.8  %Similarity = 53.8
Matches = 137  Mismatches = 175  Conservative Sub.s = 88

1749      1779      1809      1839      1869      1899      1929      1959
TRLKEADSVQQLATKRVEDKLTGLSTKLESFSVGGYNYVIDGGEPKELMANFYGKTYDINPQLLERTSQATLSFSYEAES
                                             ::                :  | :|  |      :
                                           MSRDPTYTINEHDLSFADGRFYVTFKADKSSETVRLN
                                                   10        20        30

1989      2019      2049      2079      2109      2139
TSRLEVRLYKKMHTGDTSKITIIVMPNFDLSPGKGFISQSFDLGGVMPDPRNQAWLVMRGTNANPLTL------------
:| |   : ||::  | : :   | |        |:  |  :|   : ||::  |  ::  |   |
SSCLGNTIIKKLQVEDDNTMHDFVKPKVTTQQAFGLAQQVKELDLQLKDPKSDLWGKIKFNNKAMLVEYANKEMSSAIAQ
           50        60        70        80        90       100       110

2184      2214      2244
---------------------------------SKVKLERGTVATDWNNRDETLKASFAEYKQTVDE-------------
                                   |:  |  :|: :::|   |   ::   ||:|
SAEQILLQVKSIDDERYSKFEQTLNGIKQTVKSESVESARTQLASMFDSRISGLDGKYSRLSQTIDSLSSRLDDGVGNYS
          130       140       150       160       170       180       190

2271      2301      2331      2361      2388      2418      2448
------------------NLANLRTSTETLAGQLTSAESSIRQTSESFSNRLVSLETY-KDSEPNRASRYFEASKSETAK
                  ::: |  : : |  |:|:|     |   |:|  :  |:|  :|:: |  ||
TLSQKVSGIDLRVSNAANDVSRLSQTAQGLQSQITNA----NQNYSSLSQTVQGLQTTVRDNQSNATSRI----------
          210       220       230           240       250       260

2478      2838      2868      2898      2928      2958                3009
QLSALRTEVN~~~~SFVANNANFRANSLKIRFTDSQLKFRVTTLRVTKGTIPADWSPSPDDLK-AYSDT--KLEQTANEI
                                  :||   ::| :|||| :    : | | :  |  |     : | :||
----------~~~~--------------------NQLSDLIST-KVTKGDVETTIAQSYDKIAFAIRDKLPASKMTGSEI
                                      270        280       290       300

3039      3069      3099      3129      3159      3189      3213      3243
KASVTSLDHKTLKQTDITMTSEGIVLRAGKTSNDVARAIGSYFKVTPDAIALFSSLIKVSG-NMLVDG-SVTSRKLVTGA
 :                                              || |  | :|::| |: :|| |  |  ::  |
IS----------------------------------------------AINLDRSGVKITGKNITLDGNSYISNAVIKDA
                                                       320       330       340

3261      3291      3321      3351      3381      3411      3441      3471
----VETGHVKAGAITGVLLAAEAVTAEKLKVDQAFFNKLMANDAYLKQLFAKSAFITQVQSVTISASQISGGVIKALNN
    :: | :  | :    :||||:| :|:|:| |||||| ||: |:: ||||: | | ||:|| |||:|:|||: | |
HIANMDAGKINTGYLNASRIAAEAITGDKIKMDYAFFNKLTANEGYFRTLFAKNIFTTSVQAVTTSASKITGGVLSATNG
               360       370       380       390       400       410       420

3501                3537      3567                3624      3648      3678
AMEIQMNSGQILYYTD--------QAALKRVLSGYPTQFVKFATGTVSG-KGNAGVTVIG--SNRYGTESTNDGGFVGVR
|    :||  | :  |          || |  ||  | || |:  |  | :|:|    ||  |:  | :| : | | |||
ASRWDLNSANIDFNRDATINFNSKNNALVRK-SGTNTAFVHFSNATPKGYRGSALYASIGITSSGDGIDSASSGRFCGVR
               440       450       460        470       480       490       500

3687      3717      3747      3777      3807      3837      3867      3897
-------AWNGSNIDSLDLVGDEIRLASSAFDNSDGWDVRTLDSGLKITPHNRAAERNSRIEVGDVWILKGNGSYSSLRD
         : : :|  :: ||:| : |  |:   |: :|       :  :       :                   |
FFRYAEGLQHTAKVDQAEIYGDDI-VFSDDFNIDRGFKMRPSLMPKMVDLNKMYQAILALGRCWLHANNTAWSWNFDTRS
                520        530       540       550       560       570       580
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9059> which encodes amino acid sequence <SEQ ID 9060>. An alignment of the GAS and GBS sequences follows:

```
Score = 87.8 bits (214), Expect = 4e-19
Identities = 88/273 (32%), Positives = 133/273 (48%), Gaps = 47/273 (17%)
```

```
Query: 370  AINLNSRGVQIAGKNIALDGNTT----VNGAF-------GAKLGEFI--------KLRAD 410
            AI L S  ++++G N+ +DG+ T    V GA        GA  G  +        KL+ D
Sbjct: 897  AIALFSSLIKVSG-NMLVDGSVTSRKLVTGAVETGHVKAGAITGVLLAAEAVTAEKLKVD 955

Query: 411  QIIGGTIDANKINVINLKASSIVGLDANFIKARISYAIT-DLLEGKVIKARNGAMTIDLQ 469
            Q     + AN   +  L A S       FI    S  I+   + G VIKA N AM I +
Sbjct: 956  QAFFNKLMANDAYLKQLFAKSA------FITQVQSVTISASQISGGVIKALNNAMEIQMN 1009

Query: 470  SGQINHYTNESAMRRIDSSTASQFIKMTKSGFISEIGNMQAAMTVIGSNSDGSENHENKT 529
            SGQI +YT+++A++R+ S   +QF+K   +G +S  GN  A +TVIGSN  G+E+  +
Sbjct: 1010 SGQILYYTDQAALKRVLSGYPTQFVKFA-TGTVSGKGN--AGVTVIGSNRYGTESTNDGG 1066

Query: 530  FGGIRIWNGKSSYQSTSFVELVGN--RVAIYGNKNRSPWLFDSTTSGYAYLIPQNDRGIK 587
            F G+R WNG     +   ++LVG+  R+A     N   W   +  SG   + P N
Sbjct: 1067 FVGVRAWNG----SNIDSLDLVGDEIRLASSAFDNSDGWDVRTLDSGLK-ITPHN----- 1116

Query: 588  HVIGRADRKIDQIHVGDIYV-QGERVAMMLKDL 619
                RA  +  +I VGD+++ +G      L+D+
Sbjct: 1117 ----RAAERNSRIEVGDVWILKGNGSYSSLRDI 1145
 Score = 31.3 bits (69), Expect = 0.038
 Identities = 34/151 (22%), Positives = 62/151 (40%), Gaps = 13/151 (8%)

Query: 160 QNADKKLSASYQLGIDGLKATMRSDKIGLQAEIQTTAQGLYQRYDNEIRKLSAKITTTSS 219
           Q A K  +A++       K  +  D    +A++++    L  R DN++  L+ +    +S
Sbjct: 306 QRAVKNANAAFDAEFGKTKTKINDDIEIAKAKVESFKSELSNRMDNQLLPLATEAKNLAS 365

Query: 220 GTTEAYESKLDGLRAEFTH---SNQGMRVELES--------KISGLQSTQQATARQISQE 268
                   K   LRAE      S + +++ L +        K   L   + A  R + +
Sbjct: 366 QAQADLTRKEIELRAELNRQVTSTEAVKISLTNLSHNMDIIKQKALNDLRDAETR-LKEA 424

Query: 269 ISNREGAVSRVQQGLDSYQRRLQS-AEGNYN 298
            S ++ A  RV+  L     +L+S + G YN
Sbjct: 425 DSVQQLATKRVEDKLTGLSTKLESFSVGGYN 455
```

SEQ ID 8666 (GBS202) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 5; MW 132kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 827

A DNA sequence (GBSx0877) was identified in *S.agalactiae* <SEQ ID 2503> which encodes the amino acid sequence <SEQ ID 2504>. This protein is predicted to be nuclear/mitotic apparatus protein. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2847(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 828

A DNA sequence (GBSx0879) was identified in *S.agalactiae* <SEQ ID 2505> which encodes the amino acid sequence <SEQ ID 2506>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3420(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 829

A DNA sequence (GBSx0880) was identified in *S.agalactiae* <SEQ ID 2507> which encodes the amino acid sequence <SEQ ID 2508>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.54   Transmembrane   10 -  26 (   2 -  28)

----- Final Results -----
               bacterial membrane --- Certainty=0.4015(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB07984 GB:Z93946 hypothetical protein [bacteriophage Dp-1]
 Identities = 67/136 (49%), Positives = 91/136 (66%)

Query: 1   MPPWLIDSTVVVAMVTVLGGLFSTIITTSANRKDQLIKHQYEDIKEDLSGLIDKVKTIDH 60
           MP WL D+ V+  ++T   G+ + ++      K    K   EDI   LS L  +V  ID
Sbjct: 1   MPMWLNDTAVLTTIITACSGVLTVLLNKLFEWKSNKAKSVLEDISTTLSTLKQQVDGIDQ 60

Query: 61  TTTETKKISEITKDGTLKIQRYRLFHDLTKEISQGYTTIEHFRELSILFESYQLLGGNGE 120
           TT      +++ +DGT KIQRYRL+HDL +E+  GYTT++HFRELSILFESY+ LGGNGE
Sbjct: 61  TTVAINHQNDVIQDGTRKIQRYRLYHDLKREVITGYTTLDHFRELSILFESYKNLGGNGE 120

Query: 121 IEALFEKFKQLPIEED 136
           +EAL+EK+K+LPI E+
Sbjct: 121 VEALYEKYKKLPIREE 136
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 2508 (GBS118) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 32 (lane 5; MW 42kDa).

GBS118-GST was purified as shown in Figure 198, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 830

A DNA sequence (GBSx0882) was identified in *S.agalactiae* <SEQ ID 2509> which encodes the amino acid sequence <SEQ ID 2510>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8667> and protein <SEQ ID 8668> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:      6.58
GvH: Signal Score (-7.5): -0.49
     Possible site: 53
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:  12.15 threshold:  0.0
   PERIPHERAL  Likelihood = 12.15      84
 modified ALOM score:  -2.93

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear)
```

SEQ ID 2510 (GBS56) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 8; MW 9.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 10; MW 34.9kDa).

GBS56-GST was purified as shown in Figure 195, lane 7.

Example 831

A DNA sequence (GBSx0883) was identified in *S.agalactiae* <SEQ ID 2511> which encodes the amino acid sequence <SEQ ID 2512>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 832

A DNA sequence (GBSx0884) was identified in *S.agalactiae* <SEQ ID 2513> which encodes the amino acid sequence <SEQ ID 2514>. This protein is predicted to be N-acetylmuramoyl-L-alanine amidase. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0342(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB07986 GB:Z93946 N-acetylmuramoyl-L-alanine amidase
          [bacteriophage Dp-1]
 Identities = 96/141 (68%), Positives = 118/141 (83%)

Query: 1   MEINTEIAIAWMSARQGKVSYSMDYRDGPNSYDCSSSVYYALRSAGASSAGWAVNTEYMH 60
            M ++ E  +AWM AR+G+VSYSMD+RDGP+SYDCSSS+YYALRSAGASSAGWAVNTEYMH
Sbjct: 1   MGVDIEKGVAWMQARKGRVSYSMDFRDGPDSYDCSSSMYYALRSAGASSAGWAVNTEYMH 60

Query: 61  DWLIKNGYELIAENVDWNAVRGDIAIWGMRGHSSGAGGHVVMFIDPENIIHCNWANNGIT 120
             WLI+NGYELI+EN  W+A RGDI IWG +G S+GAGGH  MFID +NIIHCN+A +GI+
Sbjct: 61  AWLIENGYELISENAPWDAKRGDIFIWGRKGASAGAGGHTGMFIDSDNIIHCNYAYDGIS 120

Query: 121 VNNYNQTAAASGWMYCYVYRL 141
            VN++++    +G  Y YVYRL
Sbjct: 121 VNDHDERWYYAGQPYYYVYRL 141
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8669> and protein <SEQ ID 8670> were also identified. Analysis of this protein sequence reveals the following:

```
RGD motif 81-83
```

The protein has homology with the following sequences in the databases:

```
58.2/72.9% over 182aa

  GP|1934766| N-acetylmuramoyl-L-alanine amidase {bacteriophage Dp-1} Insert characterized

ORF00875(301 - 1044 of 2004)
GP|1934766|emb|CAB07986.1||Z93946(1 - 183 of 296) N-acetylmuramoyl-L-alanine amidase
{bacteriophage Dp-1}
%Match = 15.5
%Identity = 58.2  %Similarity = 72.8
Matches = 107  Mismatches = 49  Conservative Sub.s = 27

234       264       294       324       354       384       414       444
LQKYNIHMSDDDLTLFVESAVKQMHDAWKE*PMEINTEIAIAWMSARQGKVSYSMDYRDXPNSYDCSSSVYYALRSAGAS
                                | :: |  :||| ||:|:||||||:|| |:|||||||:||||||||||
                                MGVDIEKGVAWMQARKGRVSYSMDFRDGPDSYDCSSSMYYALRSAGAS
                                        10        20        30        40

474       504       534       564       594       624       654       684
SAGWAVNTEYMHDWLIKNGYELIAENVDWNAVRGDIAIWGMRGHSSGAGGHVVMFIDPENIIHCNWANNGITVNNYNQTA
|||||||||||| |||:||||||:||  |:| |||| ||| :| |:|||||  |||| :||||||:| :||:||::::
SAGWAVNTEYMHAWLIENGYELISENAPWDAKRGDIFIWGRKGASAGAGGHTGMFIDSDNIIHCNYAYDGISVNDHDERW
```

```
          60        70        80        90       100       110       120

714       744       774       804       834       864       894       924
AASGWMYCYVYRLKSGASTQGKSLDTLVKETLAGNYGNGEARKAVLGNQYEAVMSVINGKTTTNQKTVDQLVQEVIAGKH
  :|  | ||||| :
YYAGQPYYYVYRLTNA----------------------------------------------------------------
         140

954       984      1014      1044      1074      1104      1134      1164
GNGEARKKSLGSQYDAVQKRVTELLKKQPSEPFKAQEVNKPTETKTSQTELTGQATATKEEGDLSFNGTILKKAVLDKIL
 | :  :| || | ||            | : |:  | ||       |  :  :      :|:  :       |   | :
-NAQPAEKKLGWQKDATGFWYARANGTYPKDEFEYIEENKSWFYFDDQGYMLAEKWLKHTDGNWYWFDRDGYMATSWKRI
              160       170       180       190       200       210       220
```

SEQ ID 8670 (GBS302) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 6; MW 55kDa).

The GBS302-His fusion product was purified (Figure 205, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 302), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 833

A DNA sequence (GBSx0885) was identified in *S.agalactiae* <SEQ ID 2515> which encodes the amino acid sequence <SEQ ID 2516>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1509(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 834

A DNA sequence (GBSx0886) was identified in *S.agalactiae* <SEQ ID 2517> which encodes the amino acid sequence <SEQ ID 2518>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1264(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB13473 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 25/68 (36%), Positives = 41/68 (59%)
```

```
Query: 4  IENLIIAIVKPLISQPDQLTIKIQDGPEFLEYHLDLDTQDIGRVIGKKGRTITAIRSIVY 63
           E+LI+ IV PL+  PD + +  ++   + +  L +   D G+VIGK+GRT  AIR+ V+
Sbjct: 6  LEDLIVHIVTPLVDHPDDIRVIREETDQKIALRLSVHKSDTGKVIGKQGRTAKAIRTAVF 65

Query: 64 SVPTQGKK 71
           +  Q  K
Sbjct: 66 AAGVQSSK 73
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2519> which encodes the amino acid sequence <SEQ ID 2520>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1012(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 72/79 (91%), Positives = 75/79 (94%)

Query: 1  MDTIENLIIAIVKPLISQPDQLTIKIQDGPEFLEYHLDLDTQDIGRVIGKKGRTITAIRS 60
          MDTIENLIIAIVKPLISQPD LTIKI+D P+FLEYHLDLD QDIGRVIGKKGRTITAIRS
Sbjct: 1  MDTIENLIIAIVKPLISQPDNLTIKIEDTPDFLEYHLDLDAQDIGRVIGKKGRTITAIRS 60

Query: 61 IVYSVPTQGKKVRLIIDEK 79
          IVYSVPT GKKVRL+IDEK
Sbjct: 61 IVYSVPTLGKKVRLVIDEK 79
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 835

A DNA sequence (GBSx0887) was identified in *S.agalactiae* <SEQ ID 2521> which encodes the amino acid sequence <SEQ ID 2522>. This protein is predicted to be ribosomal protein S16 (rpsP). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3654(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06202 GB:AP001515 ribosomal protein S16 (BS17) [Bacillus halodurans]
 Identities = 62/90 (68%), Positives = 73/90 (80%)

Query: 1  MAVKIRLTRMGSKKKPFYRINVADSRAPRDGRFIETVGTYNPLVAENQVTIKEERVLEWL 60
          MAVKIRL RMGSKK PFYR+ VADSR+PRDGRFIE +GTYNPL    +V +KE+R L+W+
Sbjct: 1  MAVKIRLKRMGSKKAPFYRVVVADSRSPRDGRFIEEIGTYNPLTQPAKVELKEDRALDWM 60

Query: 61 SKGAQPSDTVRNLLSKAGVMTKFHDQKFSK 90
           KGA+PSDTVRNL SKAG+M K H+ K  K
Sbjct: 61 LKGAKPSDTVRNLFSKAGLMEKLHNAKNEK 90
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2523> which encodes the amino acid sequence <SEQ ID 2524>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3654(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 86/90 (95%), Positives = 89/90 (98%)

Query: 1  MAVKIRLTRMGSKKKPFYRINVADSRAPRDGRFIETVGTYNPLVAENQVTIKEERVLEWL 60
          MAVKIRLTRMGSKKKPFYRINVADSRAPRDGRFIETVGTYNPLVAENQ+TIKE+RVLEWL
Sbjct: 1  MAVKIRLTRMGSKKKPFYRINVADSRAPRDGRFIETVGTYNPLVAENQITIKEDRVLEWL 60

Query: 61 SKGAQPSDTVRNLLSKAGVMTKFHDQKFSK 90
          SKGAQPSDTVRN+LSKAGVM KFHDQKFSK
Sbjct: 61 SKGAQPSDTVRNILSKAGVMAKFHDQKFSK 90
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 836

A DNA sequence (GBSx0888) was identified in *S.agalactiae* <SEQ ID 2525> which encodes the amino acid sequence <SEQ ID 2526>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.09   Transmembrane   22 -  38 (  16 -  42)
   INTEGRAL    Likelihood = -7.64   Transmembrane  382 - 398 ( 375 - 402)
   INTEGRAL    Likelihood = -7.59   Transmembrane  291 - 307 ( 284 - 317)
   INTEGRAL    Likelihood = -4.94   Transmembrane  340 - 356 ( 335 - 366)

----- Final Results -----
               bacterial membrane --- Certainty=0.5437(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC24912 GB:AF012285 YknZ [Bacillus subtilis]
 Identities = 161/417 (38%), Positives = 241/417 (57%), Gaps = 25/417 (5%)

Query: 1   MENWKFALSSILGHKMRAFLTMLGIIIGVASVVLIMALGKGMKDSVTNEITKSQKNLQIY 60
           +EN + ALSS+L HKMR+ LTMLGIIIGV SV++++A+G+G +  +   I+     +++Y
Sbjct: 4   LENIRMALSSVLAHKMRSILTMLGIIIGVGSVIVVVAVGQGGEQMLKQSISGPGNTVELY 63

Query: 61  YKTKEDQ-KNEDNFGAQGAFMQGSDTNRKEPIIQESWLKKIAKEVDGVSGYYVTNQTNAP 119
           Y   +++  +  N  A+  F +                 K  K ++G+     +   +
Sbjct: 64  YMPSDEELASNPNAAAESTFTENDI--------------KGLKGIEGIKQVVASTSESMK 109

Query: 120 VAYLEKKAKTVNITGINRTYLGIKKFKIKSGRQFQEEDYNQFSRVILLEEKLAQRLFQTN 179
             Y E++     + GIN  Y+ +   KI+SGR F + D+   +RV ++ +K+A+ LF
Sbjct: 110 ARYHEEETDAT-VNGINDGYMNVNSLKIESGRTFTDNDFLAGNRVGIISQKMAKELFDKT 168

Query: 180 EAALNKVVTVKNKSYLVVGVYSDPEAGSGLYGSNSDGNAILTNTQLASEFGAKEAENIYF 239
            + L +VV +  +   ++GV       +GL   +     +  N  + S FG  +  N+
Sbjct: 169 -SPLGEVVWINGQPVEIIGVLKKV---TGLLSFDLSEMYVPFN-MMKSSFGTSDFSNVSL 223

Query: 240 HLNDVSQSNRIGKEIGKRLTDISHAKDGYYDNFDMTSIVKSINTQVGIMTGVIGAIAAIS 299
            +         GKE  + + D +H  +  Y   +M  I   I     IMT +IG+IA IS
Sbjct: 224 QVESADDIKSAGKEAAQLVND-NHGTEDSYQVMNMEEIAAGIGKVTAIMTTIIGSIAGIS 282

Query: 300 LLVGGIGVMNIMLVSVTERTREIGLRKALGATRRKILAQFLIESMVLTILGGLIGLLLAY 359
```

```
           LLVGGIGVMNIMLVSVTERTREIG+RK+LGATR +IL QFLIES+VLT++GGL+G+ + Y
Sbjct: 283 LLVGGIGVMNIMLVSVTERTREIGIRKSLGATRGQILTQFLIESVVLTLIGGLVGIGIGY 342

Query: 360 GGTMLIANAQDKITPS-VSLNVAIGSLIFSAFIGIIFGLLPANKASKLNPIDALRYE 415
           GG  L++       PS +S  V  G ++FS  IG+IFG+LPANKA+KL+PI+ALRYE
Sbjct: 343 GGAALVSAIAG--WPSLISWQVVCGGVLFSMLIGVIFGMLPANKAAKLDPIEALRYE 397
```

There is also homology to SEQ ID 1350.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 837

A DNA sequence (GBSx0889) was identified in *S.agalactiae* <SEQ ID 2527> which encodes the amino acid sequence <SEQ ID 2528>. This protein is predicted to be ABC transporter (ATP-bindingprot). Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4080(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06841 GB:AP001517 ABC transporter (ATP-binding protein)
          [Bacillus halodurans]
 Identities = 131/218 (60%), Positives = 169/218 (77%)

Query: 8   LIRLHQIVKSYQNGDQKLQVLKNIDLTVYEGEFLAIMGPSGSGKSTLMNIIGLLDSPTSG 67
           +I+L ++ KS++ G + +++L  IDL +  G+FLAIMGPSGSGKSTLMNIIG LD PTSG
Sbjct: 1   MIKLERVTKSFRVGTEMVEILSAIDLEIASGDFLAIMGPSGSGKSTLMNIIGCLDQPTSG 60

Query: 68  DYSLNGKRVEELSQTKLAQVRNKEIGFVFQQFFLLSKLTALQNVELPLIYAGVPPKKRKN 127
            Y  +GK +   S+ ++A++RN+ IGFVFQQF LL +LTALQNVELP++YAG+  K+R
Sbjct: 61  RYMFDGKDLTNYSEQEIAKIRNRHIGFVFQQFHLLPRLTALQNVELPMVYAGMKKKERTE 120

Query: 128 LAKQFLDKVELRERMNHLPTELSGGQKQRVAIARALVNSPSIILADEPTGALDTKTGEQI 187
            A   L++V L ERM +LP  LSGGQKQRVAIAR++VN P+IILADEPTGALDTKT E I
Sbjct: 121 RAAHALERVGLAERMTYLPNSLSGGQKQRVAIARSIVNEPNIILADEPTGALDTKTSETI 180

Query: 188 MQFLTELNQEGKTIIMVTHEPEIADYATRKIVIRDGEI 225
           M+ L  LN EG TI +VTHEPEIA+Y  + + +RDG+I
Sbjct: 181 MELLCSLNNEGTTIALVTHEPEIAEYTQQTVFVRDGQI 218
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2529> which encodes the amino acid sequence <SEQ ID 2530>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1739(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 182/232 (78%), Positives = 207/232 (88%)

Query: 5   RKELIRLHQIVKSYQNGDQKLQVLKNIDLTVYEGEFLAIMGPSGSGKSTLMNIIGLLDSP 64
```

```
           +K+L++L  IVKSYQNGDQ L+VLK I+LTVYEGEFLAIMGPSGSGKSTLMNIIGLLD P
Sbjct: 5   KKQLMQLSNIVKSYQNGDQVLKVLKGINLTVYEGEFLAIMGPSGSGKSTLMNIIGLLDRP 64

Query: 65  TSGDYSLNGKRVEELSQTKLAQVRNKEIGFVFQQFFLLSKLTALQNVELPLIYAGVPPKK 124
           TSGDY+L+  ++E L+  +LA+VRN EIGFVFQQFFLL+KLTALQNVELPLIYAGV   K
Sbjct: 65  TSGDYTLHNTKIEILNDRELAKVRNDEIGFVFQQFFLLAKLTALQNVELPLIYAGVNVSK 124

Query: 125 RKNLAKQFLDKVELRERMNHLPTELSGGQKQRVAIARALVNSPSIILADEPTGALDTKTG 184
           R+  AKQFL+KV L  R+ HLP+ELSGGQKQRVAIARALVN PSIILADEPTGALDTKTG
Sbjct: 125 RREQAKQFLEKVGLGRRIKHLPSELSGGQKQRVAIARALVNDPSIILADEPTGALDTKTG 184

Query: 185 EQIMQFLTELNQEGKTIIMVTHEPEIADYATRKIVIRDGEITADTTDSIRID 236
           +QIM+ LTELN+EGKTIIMVTHEPEIAD+ATRKI+IRDG+IT DTT S+ ID
Sbjct: 185 QQIMELLTELNKEGKTIIMVTHEPEIADFATRKIIIRDGDITTDTTASVVID 236
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 838

A DNA sequence (GBSx0890) was identified in *S.agalactiae* <SEQ ID 2531> which encodes the amino acid sequence <SEQ ID 2532>. This protein is predicted to be ATP-binding cassette transporter-like protein. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.97   Transmembrane   17 -  33 (  13 -  39)

----- Final Results -----
               bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9965> which encodes amino acid sequence <SEQ ID 9966> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC24909 GB:AF012285 YknX [Bacillus subtilis]
 Identities = 104/391 (26%), Positives = 182/391 (45%), Gaps = 21/391 (5%)

Query: 13  KKGAIISGLSVALIVVIGGFLWVQSQPNKSAVKTNYKVFNVREGSVSSSTLLTGKAKANQ 72
           KK  I  G++V + + +G  ++  + P   +     +  +V E  +SS+ ++ G  K +
Sbjct: 2   KKVWIGIGIAVIVALFVGINIYRSAAPTSGSAGKEVQTGSVEENEISSTVMVPGTLKFSN 61

Query: 73  EQYVYFDANKGNRATVTVKVGDKITAGQQLVQYDTTTAQAAYDTANRQLNKVARQINNLK 132
           EQYV+++A+KG    + VK GDK+  G  LV Y  T  Q + +    QL   + ++   +
Sbjct: 62  EQYVFYEADKGTLEDIKVKEGDKVKKGTALVTY--TNEQLSLEKEQNQLTSESNRLQIDQ 119

Query: 133 TTGSLPAMESSDQSSSSSQGQGTQSTSGATNRLQQNYQSQANASYNQQLQDLNDAYADAQ 192
               L A++S ++      G+        + R +   Q +          +L       Q
Sbjct: 120 IQEKLKALDSKERELEKQVGKKEAEKQIESERTELQMQKKTAEI------ELKQTELQRQ 173

Query: 193 AEVNKAQKALNDTVITSDVSGTVVEVNSDIDPASKTSQV---LVHVATEGKLQVQGTMSE 249
           +  N+    ++D  + S++ GTV+ VN   + ASK S +   ++H+     L V G +SE
Sbjct: 174 SLANR----VSDLEVKSEIEGTVISVNQ--EAASKKSDIQEPVIHIGNPKDLVVSGKLSE 227

Query: 250 YDLANVKKDQAVKIKSKVYPDKEWEGKISYISNYPEAEANNNDSNNGSSAVNYKYKVDIT 309
           YD   VKK Q V + S V   K W+G +S +   P+ +  +  +     AV Y  +V I
Sbjct: 228 YDTLKVKKGQKVTLTSDVIQGKTWKGTVSAVGLVPD-QQESAAAQGTEQAVQYPLQVKIK 286

Query: 310 SPLDALKQGFTVSVEV-VNGDKHLIVPTSSVINKDNKHFVWVYNDSNRKISKVEVKIGKA 368
             L   K GF   + +  +  K   +P+ +V  +D++++V+   D   K  +V+VKIG+
Sbjct: 287 GNLPEGKPGFKFIMNIETDKRKANTLPSKAVKKEDDQYYVYTVKDG--KAKRVDVKIGEV 344
```

```
Query: 369 DAKTQEILSGLKAGQIVVTNPSKTFKDGQKI 399
                EI  GL     V+ NPS    DG ++
Sbjct: 345 TDDLTEIKEGLTQDDQVILNPSDQVTDGMEV 375
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2533> which encodes the amino acid sequence <SEQ ID 2534>. Analysis of this protein sequence reveals the following:

```
     Possible site: 42
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.61    Transmembrane   15 -  31 (  11 -  36)

----- Final Results -----
               bacterial membrane --- Certainty=0.4843(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC24909 GB:AF012285 YknX [Bacillus subtilis]
 Identities = 103/380 (27%), Positives = 180/380 (47%), Gaps = 21/380 (5%)

Query: 16  ITASVITLVLIITGIVLWKQQRNTLTADIAKEPYSTVSVTEGSIASSTLLSGTVKALSEE 75
            I   +  +V +  GI +++    T  +  A +   T SV E  I+S+ ++ GT+K  +E+
Sbjct: 6   IGIGIAVIVALFVGINIYRSAAPT--SGSAGKEVQTGSVEENEISSTVMVPGTLKFSNEQ 63

Query: 76  YIYFDANKGNDATVTVKVGDQVTQGQQLVQYNTTTAQSAYDTAVRSLNKIGRQINHLKTY 135
            Y++++A+KG    + VK GD+V +G  LV Y  T  Q + +   +  N++  + N L+
Sbjct: 64  YVFYEADKGTLEDIKVKEGDKVKKGTALVTY--TNEQLSLE---KEQNQLTSESNRLQID 118

Query: 136 GVPAVSTETNRDEATGEETTTTVQPSAQ-QNANYKQQLQDLNDAYADAQAEVNKAQIA-- 192
             +       +  E   E+     +   Q ++   + Q+Q         Q E+ +  +A
Sbjct: 119 QIQEKLKALDSKERELEKQVGKKEAEKQIESERTELQMQKKTAEIELKQTELQRQSLANR 178

Query: 193 LNDTVVISSVSGTVVEVNND-IDPSSKNSQTLVHVATEGQLQVKGTLTEYDLANVKVGQS 251
            ++D  V S + GTV+ VN +     S   + ++H+     L V G L+EYD   VK GQ
Sbjct: 179 VSDLEVKSEIEGTVISVNQEAASKKSDIQEPVIHIGNPKDLVVSGKLSEYDTLKVKKGQK 238

Query: 252 VKIKSKVYSNQEWTGKISYVSNYPTESNAGSTTPAGSTGAGSSTGATYDYKIDIISPLNQ 311
            V + S V   + W G +S V   P +  +       +   G+     Y  ++ I   L +
Sbjct: 239 VTLTSDVIQGKTWKGTVSAVGLVPDQQES-------AAAQGTEQAVQYPLQVKIKGNLPE 291

Query: 312 LKQGFTVSVEVVNEAKQA-LVPLTAVIKKDKKHYVWTYDDATGKAKKVEVTLGNADAQQQ 370
             K GF   + +  + ++A  +P  AV K+D ++YV+T  D  GKAK+V+V +G
Sbjct: 292 GKPGFKFIMNIETDKRKANTLPSKAVKKEDDQYYVYTVKD--GKAKRVDVKIGEVTDDLT 349

Query: 371 EIHKGVAVGDIVIANPDKNI 390
            EI +G+   D VI NP   +
Sbjct: 350 EIKEGLTQDDQVILNPSDQV 369
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 234/421 (55%), Positives = 301/421 (70%), Gaps = 19/421 (4%)

Query: 3   MSKRQNLGISKKGAIISGLSVALIVVIGGF-LWVQSQPNKSA--VKTNYKVFNVREGSVS 59
            MSKR  + I+ K  +I+   + L+++I G  LW Q +   +A   K  Y   +V EGS++
Sbjct: 1   MSKRGKIKITTKTKLITASVITLVLIITGIVLWKQQRNTLTADIAKEPYSTVSVTEGSIA 60

Query: 60  SSTLLTGKAKANQEQYVYFDANKGNRATVTVKVGDKITAGQQLVQYDTTTAQAAYDTANR 119
            SSTLL+G  KA  E+Y+YFDANKGN ATVTVKVGD++T GQQLVQY+TTTAQ+AYDTA R
Sbjct: 61  SSTLLSGTVKALSEEYIYFDANKGNDATVTVKVGDQVTQGQQLVQYNTTTAQSAYDTAVR 120

Query: 120 QLNKVARQINNLKTTGSLPAMESSDQSSSSSQGQGTQSTSGATNRLQQNYQSQANASYNQ 179
             LNK+ RQIN+LKT G +PA+ S++ +   + G+ T +T   +        +Q NA+Y Q
Sbjct: 121 SLNKIGRQINHLKTYG-VPAV-STETNRDEATGEETTTTVQPS--------AQQNANYKQ 170

Query: 180 QLQDLNDAYADAQAEVNKAQKALNDTVITSDVSGTVVEVNSDIDPASKTSQVLVHVATEG 239
```

```
           QLQDLNDAYADAQAEVNKAQ ALNDTV+ S VSGTVVEVN+DIDP+SK SQ LVHVATEG
Sbjct: 171 QLQDLNDAYADAQAEVNKAQIALNDTVVISSVSGTVVEVNNDIDPSSKNSQTLVHVATEG 230

Query: 240 KLQVQGTMSEYDLANVKKDQAVKIKSKVYPDKEWEGKISYISNYP-EAEANN-----NDS 293
           +LQV+GT++EYDLANVK  Q+VKIKSKVY ++EW GKISY+SNYP E+ A +      +
Sbjct: 231 QLQVKGTLTEYDLANVKVGQSVKIKSKVYSNQEWTGKISYVSNYPTESNAGSTTPAGSTG 290

Query: 294 NNGSSAVNYKYKVDITSPLDALKQGFTVSVEVVNGDKHLIVPTSSVINKDNKHFVWVYND 353
              S+   Y YK+DI SPL+ LKQGFTVSVEVVN  K  +VP ++VI KD KH+VW Y+D
Sbjct: 291 AGSSTGATYDYKIDIISPLNQLKQGFTVSVEVVNEAKQALVPLTAVIKKDKKHYVWTYDD 350

Query: 354 SNRKISKVEVKIGKADAKTQEILSGLKAGQIVVTNPSKTFKDGQKIDNIESIDLNSNKKSE 414
           +  K  KVEV +G ADA+ QEI  G+  G IV+ NP K  K  +K++ + SI  N+  + +
Sbjct: 351 ATGKAKKVEVTLGNADAQQQEIHKGVAVGDIVIANPDKNIKPDKKLEGVISIGTNTKPEKD 411
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 839

A DNA sequence (GBSx0891) was identified in *S.agalactiae* <SEQ ID 2535> which encodes the amino acid sequence <SEQ ID 2536>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1832(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 840

A DNA sequence (GBSx0892) was identified in *S.agalactiae* <SEQ ID 2537> which encodes the amino acid sequence <SEQ ID 2538>. This protein is predicted to be carbamoyl-phosphate synthase, pyrimidine-specific, large chain, putati. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.70   Transmembrane  486 - 502 ( 486 - 502)

----- Final Results -----
               bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA91005 GB:Z54240 carbamoyl-phosphate synthase [Lactobacillus
           plantarum]
 Identities = 117/417 (28%), Positives = 205/417 (49%), Gaps = 37/417 (8%)

Query: 122 FVQVDCLVMRDSLNNCLYVSDLEYIES-NKTTGKSLAIVPSQTLSDAARQTIRDVAFDVC 180
           + +++  VMRD+ +N + V ++E  +     TG S+   P QTL+D   Q +RD A  +
Sbjct: 213 YKEIEFEVMRDAADNAMVVCNMENFDPVGIHTGDSIVYAPVQTLADREVQLLRDAALKII 272
```

```
Query: 181 RKANIIGVCYFSFLIDLNSLDYHIISLSSGLSHQSILFETITTYPVLEIATKLTVGYTFS 240
           R   I G C     +D NS +Y+II ++  +S  S L    T YP+ ++A K+ VG
Sbjct: 273 RALKIEGGCNVQLALDPNSFNYYIIEVNPRVSRSSALASKATGYPIAKMAAKIAVGLHLD 332

Query: 241 QLKHSYYPNTSAFLEPQLDYVATV--SFSFEKVDY---------------IFFARNIEQL 283
           ++K+     T A  EP LDYV      + F+K  +               +   RNIE+
Sbjct: 333 EIKNPVTGTTYAEFEPALDYVVCKIPRWPFDKFTHADRRLGTQMKATGEVMAIGRNIEEA 392

Query: 284 FLNLLEASS----HDHFPFLSDISEEDLMFALIQKKENRLAYLLEAFRRGFDLYDLSSVT 339
            L  + +      H     L  + ++ L   LI  +++RL YL EA RRG+ + +L+ +T
Sbjct: 393 TLKAVRSLEIGVHHVEESTLRSVDDDVLSDKLIHAQDDRLFYLTEAIRRGYQIDELAELT 452

Query: 340 KINPFYLDKCLHIVELYENLNKSQYNVDIYKEAKRYGFSDDYIASSWQISLIDMLEYRKK 399
           KIN F+LDK LHI+E+ + L     +++    AKR GF+D  +A  W  ++  + ++R
Sbjct: 453 KINVFFLDKLLHIIEIEQALRTHTDDIETLTVAKRNGFADQTVADYWHETIDQVRDFRLA 512

Query: 400 HSVAPVLKQVEQSSGVLTGHQIQYFRSYDWHSDYISSGCQKALIM----------VDKGY 449
           H +APV K V+  +G        Y+ +Y++ ++ I +     L++          V+  Y
Sbjct: 513 HKLAPVYKMVDTCAGEFASETPYYYGTYEFENESIVTKRPSVLVLGSGPIRIGQGVEFDY 572

Query: 450 SLVKLNELIKQIKQTHLELLIVTNQPLLIEQLNDTS--IIFDTIGIETILTIMGIEE 504
           + V     +K I++   E +I+ + P  +      S  + F+ + IE +L ++ +E+
Sbjct: 573 ATV---HSVKAIQKAGYEAIIMNSNPETVSTDFSVSDKLYFEPLTIEDVLNVIELEK 626
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 841

A DNA sequence (GBSx0893) was identified in *S.agalactiae* <SEQ ID 2539> which encodes the amino acid sequence <SEQ ID 2540>. This protein is predicted to be carbamoyl phosphate synthetase small subunit (carA). Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2709(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB89872 GB:AJ132624 carbamoyl phosphate synthetase small
          subunit [Lactococcus lactis]
 Identities = 188/352 (53%), Positives = 265/352 (74%)

Query: 1   MAKKLLILEDGTVFEGLSFGSSLDVTGELVFCTGNTGYQEIITNPSHNGKILVFTSPLIG 60
           M+K+LLILEDGT+FEG + G++LDVTGELVF TG TGYQE IT+ S+NG+IL FT P++G
Sbjct: 1   MSKRLLILEDGTIFEGEALGANLDVTGELVFNTGMTGYQESITDQSYNGQILTFTYPIVG 60

Query: 61  NYGIHRSYSEAIIPTCLGVVVAEYSRCVSSDTSKMNLDEFLKMKKVPAMSGVDTRYLMQV 120
           NYG++R   E+I PTC  VVV E +R  S+   +M+ DEFLK K +P ++GVDTR + ++
Sbjct: 61  NYGVNRDDYESIHPTCKAVVVHEAARRPSNWRMQMSFDEFLKSKNIPGITGVDTRAITKI 120

Query: 121 IKEKGFVKATLAEAGDVLSHLQDQLIATVLPTNNVEQVSTKTAYPSPASGRNIVVLDFGL 180
           ++E G +KA+L +A D + H   QL ATVLPTN VE  ST TAYPSP +GR +VV+DFGL
Sbjct: 121 VREHGTMKASLVQARDEVDHQMSQLQATVLPTNQVETSSTATAYPSPNTGRKVVVVDFGL 180

Query: 181 KHSILRELSKRQCDVTVIPYNTSLEGIKNLYPEGIILSNGPGNPEKLQEILNTIKELQKS 240
           KHSILRELSKR+C++TV+PYNTS + I  + P+G++L+NGPG+P  + E +  IKE+Q
Sbjct: 181 KHSILRELSKRECNLTVVPYNTSAKEILEMEPDGVMLTNGPGDPTDVPEAIEMIKEVQGK 240

Query: 241 VPMLGIGLGHQLIAMANGAEIMRLPVAKKGPNYPMRDIATGRLETVSQFNHFTVNRLNLP 300
           +P+ GI LGHQL ++ANGA   ++    +G N+ +R++ATGR++  SQ + + V+  NLP
```

```
Sbjct: 241 IPIFGICLGHQLFSLANGATTYKMKFGHRGFNHAVREVATGRIDFTSQNHGYAVSSENLP 300

Query: 301 HDLLVTHEGLNDQEIVALRHRSFPVMSVQFYPEAAPGPHDVTYFFDEFLEMI 352
            DL++TH  +ND  +  +RH+ FP  SVQF+P+AAPGPHD +Y FD+F++++
Sbjct: 301 EDLMITHVEINDNSVEGVRHKYFPAFSVQFHPDAAPGPHDASYLFDDFMDLM 352
```

There is also homology to SEQ ID 2030.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 842

A DNA sequence (GBSx0894) was identified in *S.agalactiae* <SEQ ID 2541> which encodes the amino acid sequence <SEQ ID 2542>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3646(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9967> which encodes amino acid sequence <SEQ ID 9968> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB89869 GB:AJ132624 pyrimidine regulatory protein [Lactococcus
           lactis]
 Identities = 127/169 (75%), Positives = 147/169 (86%)

Query: 13  MKRKEIIDDVTMKRAITRITYEIIERNKNLDNIVLAGIKTRGVFLAKRIQERLKQLENLD 72
           M RKEIID++TMKRAITRITYEIIERNK LD +VL GIKTRGV+LAKRIQERL+QLE L+
Sbjct: 1   MARKEIIDEITMKRAITRITYEIIERNKELDKLVLIGIKTRGVYLAKRIQERLQQLEGLE 60

Query: 73  IPVGELDTKPFRDDMKVEVDTTTMPVDITDKDIILIDDVLYTGRTIRAAIDNLVSLGRPS 132
           IP GELDT+PFRDD + + DTT + +DIT KD+IL+DDVLYTGRTIRAAID +V LGRP+
Sbjct: 61  IPFGELDTRPFRDDKQAQEDTTEIDIDITGKDVILVDDVLYTGRTIRAAIDGIVKLGRPA 120

Query: 133 RVSLAVLIDRGHRELPIRADYVGKNIPTSQFEEILVEVMEHDGYDRVSI 181
           RV LAVL+DRGHRELPIRADYVGKNIPT   EEI+V++ EHDG D + I
Sbjct: 121 RVQLAVLVDRGHRELPIRADYVGKNIPTGHDEEIIVQMSEHDGNDSILI 169
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2543> which encodes the amino acid sequence <SEQ ID 2544>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3870(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 147/171 (85%), Positives = 158/171 (91%)

Query: 13  MKRKEIIDDVTMKRAITRITYEIIERNKNLDNIVLAGIKTRGVFLAKRIQERLKQLENLD 72
           MK KEI+DDVTMKRAITRITYEIIERNK LDN+VLAGIKTRGVFLA+RIQERL QLE LD
Sbjct: 1   MKTKEIVDDVTMKRAITRITYEIIERNKQLDNVVLAGIKTRGVFLARRIQERLHQLEGLD 60
```

```
Query: 73  IPVGELDTKPFRDDMKVEVDTTTMPVDITDKDIILIDDVLYTGRTIRAAIDNLVSLGRPS 132
           +P+GELD KPFRDDM+VE DTT M VDIT KD+ILIDDVLYTGRTIRAAIDNLVSLGRP+
Sbjct: 61  LPIGELDIKPFRDDMRVEEDTTLMSVDITGKDVILIDDVLYTGRTIRAAIDNLVSLGRPA 120

Query: 133 RVSLAVLIDRGHRELPIRADYVGKNIPTSQFEEILVEVMEHDGYDRVSIID 183
           RVSLAVL+DRGHRELPIRADYVGKNIPTS  EEI+VEV+E DG DRVSIID
Sbjct: 121 RVSLAVLVDRGHRELPIRADYVGKNIPTSSVEEIVVEVVEVDGRDRVSIID 171
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 843

A DNA sequence (GBSx0895) was identified in *S.agalactiae* <SEQ ID 2545> which encodes the amino acid sequence <SEQ ID 2546> (rluD). Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0687(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9969> which encodes amino acid sequence <SEQ ID 9970> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06261 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 178/290 (61%), Positives = 216/290 (74%), Gaps = 2/290 (0%)

Query: 17  GVRLDKAL-ADNSELSRSQANEEIKKGIVLVNGQVKKAKYTVQEGDRITFDIPKEEVLDY 75
           G R+DK L A   E SR+Q  + IK G VL+NG+  K+ Y V+ GD +   +P+ EVL+
Sbjct: 15  GERIDKFLTAQGEEWSRTQVQQWIKDGHVLINGRTIKSNYKVETGDTLELFVPEPEVLEV 74

Query: 76  QAENIPLDIIYQDDDVAVVNKPQGMVVHPSAGHSSGTLVNALMYHIKDLSSINGVVRPGI 135
             ENIP++IIY+D+DVAVVNKP+GMVVHP+ GH++GTLVNALMYH  DLSSINGVVRPGI
Sbjct: 75  VPENIPIEIIYEDEDVAVVNKPRGMVVHPAPGHTTGTLVNALMYHCNDLSSINGVVRPGI 134

Query: 136 VHRIDKDTSGLLMVAKNDRAHQVLAEELKDKKSLRKYLAIVHGNLPNDRGVIEAPIGRSD 195
           VHRIDKDTSGLLM+AKNDRAH+ L  +LK K + R Y AIVHGN+P+D G I+APIGR
Sbjct: 135 VHRIDKDTSGLLMIAKNDRAHESLVNQLKAKTTERVYQAIVHGNIPHDHGTIDAPIGRDK 194

Query: 196 KDRKKQAVTAK-GKPAITRFHVLERFGDYTLVELSLETGRTHQIRVHMAYIGHPLAGDPV 254
            DR+   VT +  + A+T F VLERFGD+T VE  LETGRTHQIRVH  YIG PLAGDP
Sbjct: 195 VDRQSMTVTEENSRDAVTHFTVLERFGDFTFVECQLETGRTHQIRVHFKYIGFPLAGDPK 254

Query: 255 YGPRKTLGGKGQFLHAQTLGFTHPSNGENLIFSVEVPEIFQTTLEKLRKN 304
           YGP+KTL   GQ LHAQ LGF HP  GE + F VE+PE  +  + +L+ N
Sbjct: 255 YGPKKTLSIDGQALHAQKLGFEHPRTGEFMRFKVEMPEEMKKLIRQLQNN 304
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2547> which encodes the amino acid sequence <SEQ ID 2548>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2455(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 239/295 (81%), Positives = 265/295 (89%)

Query: 9   MEITIKIAGVRLDKALADNSELSRSQANEEIKKGIVLVNGQVKKAKYTVQEGDRITFDIP 68
           MEI +  +G RLDKALAD S LSR QAN++IK+G+VLVNGQ KKAKYTVQ GD I F++P
Sbjct: 1   MEINVITSGQRLDKALADLSPLSRGQANDQIKQGLVLVNGQQKKAKYTVQAGDVICFELP 60

Query: 69  KEEVLDYQAENIPLDIIYQDDDVAVVNKPQGMVVHPSAGHSSGTLVNALMYHIKDLSSIN 128
           KEEVL+YQA+NIPLDIIY+DD +A++NKPQGMVVHPSAGH SGT+VNALMYHIKDLSSIN
Sbjct: 61  KEEVLEYQAQNIPLDIIYEDDALAIINKPQGMVVHPSAGHPSGTMVNALMYHIKDLSSIN 120

Query: 129 GVVRPGIVHRIDKDTSGLLMVAKNDRAHQVLAEELKDKKSLRKYLAIVHGNLPNDRGVIE 188
           GVVRPGIVHRIDKDTSGLLMVAK D AHQ LAEELK KKSLRKYLAIVHGNLPNDRG+IE
Sbjct: 121 GVVRPGIVHRIDKDTSGLLMVAKTDAAHQALAEELKAKKSLRKYLAIVHGNLPNDRGMIE 180

Query: 189 APIGRSDKDRKKQAVTAKGKPAITRFHVLERFGDYTLVELSLETGRTHQIRVHMAYIGHP 248
           APIGRS+KDRKKQAVTAKGK A+TRF VLERFGDY+LVEL LETGRTHQIRVHMAYIGHP
Sbjct: 181 APIGRSEKDRKKQAVTAKGKEAVTRFTVLERFGDYSLVELQLETGRTHQIRVHMAYIGHP 240

Query: 249 LAGDPVYGPRKTLGGKGQFLHAQTLGFTHPSNGENLIFSVEVPEIFQTTLEKLRK 303
           +AGDP+YGPRKTL G GQFLHA+TLG THP  G+ +IF+VE PEIFQ  L+ LRK
Sbjct: 241 VAGDPLYGPRKTLSGHGQFLHAKTLGLTHPMTGKEMIFTVEAPEIFQKVLKLLRK 295
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 844

A DNA sequence (GBSx0896) was identified in *S.agalactiae* <SEQ ID 2549> which encodes the amino acid sequence <SEQ ID 2550>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0496(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD53064 GB:AF163833 CpsY [Streptococcus agalactiae]
 Identities = 105/297 (35%), Positives = 163/297 (54%), Gaps = 4/297 (1%)

Query: 1   MNIQQLRYVVAIANSGTFREAAAKLFVSQPSLSVAVRDLETELGFQIFTRTTTGAVLTNQ 60
           M IQQL+YV+ I  +G+  EAA +L+++QPSLS AVR+LETE+G QIF R   G  LT
Sbjct: 1   MRIQQLQYVIKIVETGSMNEAAKQLYITQPSLSNAVRNLETEMGIQIFIRNPKGITLTKD 60

Query: 61  GMTFYENALEVVKSFDSFEKQFSQSEATEQEFSIASQHYDFLPPLITAFSKCNDNFSY-F 119
           GM F   A ++++     E+++     + + FS++SQHY F+     A     D   Y
Sbjct: 61  GMEFLSYARQILEQTALLEERYKGDNTSRELFSVSSQHYAFVVNAFVALFNGTDMTQYEL 120

Query: 120 RIFESTTIRILDEVAQGNSEIGIIYINSQNKKGLLQRLDKLGLEFVELIPFKTHIYLGKD 179
            + E+ T  I+D+V    SEIG++++NS N+  L +  D   L    L     HI++ K
Sbjct: 121 FLRETRTWEIIDDVKNFRSEIGVLFLNSYNRDVLTKLFDDNSLIATTLFTTTPHIFVSKS 180

Query: 180 HPLASKTSLIMTDLEGLPTVRFTQDRDDYRYYSENFVEVLDSSVTYNVTDRATLNGILER 239
           +PLA++  L M DLE  P + + Q   +  Y+SE  +  +    +  V+DRATL  ++
Sbjct: 181 NPLANRKKLSMKDLEDYPYLSYDQGLHNSFYFSEEMMSQIPHPKSIVVSDRATLFNLMIG 240

Query: 240 TQAYATGSGFLDSRSVNG--ITVIPLEDHLDNQMIYIKRKDRNLSQMALKFVAVMEE 294
              Y   +G L+S+ +NG  I  IPL+      ++YI+    NLS+M  KF+  + E
Sbjct: 241 LDGYTVATGILNSK-LNGDEIVAIPLDVDDVIDIVYIRHDKANLSKMGQKFIDYLLE 296
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2551> which encodes the amino acid sequence <SEQ ID 2552>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1252(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 217/296 (73%), Positives = 253/296 (85%)

Query: 1   MNIQQLRYVVAIANSGTFREAAAKLFVSQPSLSVAVRDLETELGFQIFTRTTTGAVLTNQ 60
            MNIQQLRYVVAIAN+GTFREAA+KLFVSQPSLSV+++DLE ELGFQIF RTT+G VLT+Q
Sbjct: 1   MNIQQLRYVVAIANNGTFREAASKLFVSQPSLSVSIKDLEAELGFQIFNRTTSGTVLTSQ 60

Query: 61  GMTFYENALEVVKSFDSFEKQFSQSEATEQEFSIASQHYDFLPPLITAFSKCNDNFSYFR 120
            G+ FYE ALEVVKSFDSFEK FSQ++  + EFSIASQHYDFLPPLITAFS+  D    FR
Sbjct: 61  GLVFYEKALEVVKSFDSFEKTFSQADLDQNEFSIASQHYDFLPPLITAFSQQYDGHRVFR 120

Query: 121 IFESTTIRILDEVAQGNSEIGIIYINSQNKKGLLQRLDKLGLEFVELIPFKTHIYLGKDH 180
            IFESTTI+ILDEVAQGNSEIGIIY+N  N+KGL QR+DKLGLE+V LIPF THIYL K H
Sbjct: 121 IFESTTIQILDEVAQGNSEIGIIYLNVDNQKGLFQRMDKLGLEYVSLIPFTTHIYLSKTH 180

Query: 181 PLASKTSLIMTDLEGLPTVRFTQDRDDYRYYSENFVEVLDSSVTYNVTDRATLNGILERT 240
            PLA++ +L + D++GLP VRFTQ+RD+Y YYSENFV+  +    YNV+DRATLNGILERT
Sbjct: 181 PLANREALYLNDIQGLPAVRFTQERDEYLYYSENFVDTSECPRIYNVSDRATLNGILERT 240

Query: 241 QAYATGSGFLDSRSVNGITVIPLEDHLDNQMIYIKRKDRNLSQMALKFVAVMEEYF 296
             A+ATGSGFLD RSVNGI VIPL DH+DNQMIY+KRKD+NLS     FV ++++YF
Sbjct: 241 NAFATGSGFLDHRSVNGIKVIPLADHIDNQMIYVKRKDKNLSVAGATFVTILKDYF 296
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 845

A DNA sequence (GBSx0897) was identified in *S.agalactiae* <SEQ ID 2553> which encodes the amino acid sequence <SEQ ID 2554>. This protein is predicted to be 50S ribosomal protein L27 (rpmA). Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0976(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14754 GB:Z99118 ribosomal protein L27 (BL24) [Bacillus subtilis]
 Identities = 70/90 (77%), Positives = 80/90 (88%)

Query: 8  NLQLFAHKKGGGSTSNGRDSQAKRLGAKAADGQTVSGGSILYRQRGTHIYPGANVGRGGD 67
           +LQ FA KKG GST NGRDS+AKRLGAK ADGQ V+GGSILYRQRGT IYPG NVGRGGD
Sbjct: 5  DLQFFASKKGVGSTKNGRDSEAKRLGAKRADGQFVTGGSILYRQRGTKIYPGENVGRGGD 64

Query: 68 DTLFAKVEGVVRFERKGRDKKQVSVYPIAK 97
           DTLFAK++G V+FER GRD+K+VSVYP+A+
Sbjct: 65 DTLFAKIDGTVKFERFGRDRKKVSVYPVAQ 94
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2555> which encodes the amino acid sequence <SEQ ID 2556>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0976(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 95/97 (97%), Positives = 96/97 (98%)

Query: 1  MLKMNLANLQLFAHKKGGGSTSNGRDSQAKRLGAKAADGQTVSGGSILYRQRGTHIYPGA 60
          MLKMNLANLQLFAHKKGGGSTSNGRDSQAKRLGAKAADGQTVSGGSILYRQRGTHIYPG
Sbjct: 1  MLKMNLANLQLFAHKKGGGSTSNGRDSQAKRLGAKAADGQTVSGGSILYRQRGTHIYPGV 60

Query: 61 NVGRGGDDTLFAKVEGVVRFERKGRDKKQVSVYPIAK 97
          NVGRGGDDTLFAKVEGVVRFERKGRDKKQVSVYP+AK
Sbjct: 61 NVGRGGDDTLFAKVEGVVRFERKGRDKKQVSVYPVAK 97
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 846

A DNA sequence (GBSx0898) was identified in *S.agalactiae* <SEQ ID 2557> which encodes the amino acid sequence <SEQ ID 2558>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.75   Transmembrane   32 -  48 (  32 -  48)

----- Final Results -----
               bacterial membrane --- Certainty=0.1298(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06729 GB:AP001517 unknown conserved protein in B. subtilis
           [Bacillus halodurans]
 Identities = 33/107 (30%), Positives = 63/107 (58%), Gaps = 4/107 (3%)

Query: 1   MIKATFTRNQSGYLYSAEISGHAGSGEYGFDVICAAVSTLSINFINSLEALTTCQAQLII 60
            MI   F RN+   + S  +SGHA +G YG D++CA  S +++  +N++ AL  CQ +L+
Sbjct: 1   MIDVVFERNKQNDIVSFTMSGHADAGPYGQDLVCAGASAVALGTVNAIIAL--CQVELVT 58

Query: 61  N-DVEGGYMKIDL-SSIPQHKEDKVQLLFESYLLGMTNLSKDSSEFV 105
              + EGG+++  + + + +   +KVQLL E   + + ++++   E +
Sbjct: 59  EMENEGGFLRCRVPNDLEETTFEKVQLLLEGMNISLQSIAESYGEHI 105
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2559> which encodes the amino acid sequence <SEQ ID 2560>. Analysis of this protein sequence reveals the following:

```
     Possible site: 52
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.59   Transmembrane   32 -  48 (  32 -  48)

----- Final Results -----
               bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB06729 GB:AP001517 unknown conserved protein in B. subtilis
          [Bacillus halodurans]
 Identities = 33/109 (30%), Positives = 60/109 (54%), Gaps = 4/109 (3%)

Query: 1   MIKAIFTRQKNGQLSSVTLTGHAGSGKHGFDIVCASVSTLAINFVNSLEVLADCQALVDL 60
           MI  +F R K   + S T++GHA +G +G D+VCA  S +A+  VN++  L   + + ++
Sbjct: 1   MIDVVFERNKQNDIVSFTMSGHADAGPYGQDLVCAGASAVALGTVNAIIALCQVELVTEM 60

Query: 61  NDVEGGYMAITIP---PHDNKEEVQLLFESFLLGMTSLAKDSSKFVNTQ 106
            + EGG++   +P        E+VQLL E   + + S+A+   + +  +
Sbjct: 61  EN-EGGFLRCRVPNDLEETTFEKVQLLLEGMNISLQSIAESYGEHIQIE 108
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 67/110 (60%), Positives = 90/110 (80%), Gaps = 2/110 (1%)

Query: 1   MIKATFTRNQSGYLYSAEISGHAGSGEYGFDVICAAVSTLSINFINSLEALTTCQAQLII 60
           MIKA FTR ++G L S  ++GHAGSG++GFD++CA+VSTL+INF+NSLE L  CQA + +
Sbjct: 1   MIKAIFTRQKNGQLSSVTLTGHAGSGKHGFDIVCASVSTLAINFVNSLEVLADCQALVDL 60

Query: 61  NDVEGGYMKIDLSSIPQHKEDKVQLLFESYLLGMTNLSKDSSEFVSTVVM 110
           NDVEGGYM I +   P   +++VQLLFES+LLGMT+L+KDSS+FV+T V+
Sbjct: 61  NDVEGGYMAITIP--PHDNKEEVQLLFESFLLGMTSLAKDSSKFVNTQVI 108
```

SEQ ID 2558 (GBS433) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 78 (lane 4; MW 16kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 8; MW 41kDa).

GBS433-GST was purified as shown in Figure 223, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 847

A DNA sequence (GBSx0899) was identified in *S.agalactiae* <SEQ ID 2561> which encodes the amino acid sequence <SEQ ID 2562>. This protein is predicted to be ribosomal protein L21 (rplU). Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2972(Affirmative) < succ>
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14756 GB:Z99118 ribosomal protein L21 (BL20) [Bacillus subtilis]
 Identities = 67/101 (66%), Positives = 78/101 (76%)

Query: 4   YAIIKTGGKQVKVEVGQAIYVEKLDVEAGAEVTFNEVVLVGGETTKVGTPVVEGATVVGT 63
           YAIIKTGGKQ+KVE GQ +Y+EKL  EAG  VTF +V+ VGG+  KVG P VEGATV
Sbjct: 2   YAIIKTGGKQIKVEEGQTVYIEKLAAEAGETVTFEDVLFVGGDNVKVGNPTVEGATVTAK 61

Query: 64  VEKQGKQKKVVSYKYKPKKGSHRKQGHRQPYTKVVINAINA 104
           VEKQG+ KK+  ++YKPKK  H+KQGHRQPYTKV I  INA
Sbjct: 62  VEKQGRAKKITVFRYKPKKNVHKKQGHRQPYTKVTIEKINA 102
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2563> which encodes the amino acid sequence <SEQ ID 2564>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3026(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 97/104 (93%), Positives = 101/104 (96%)

Query: 1   MSTYAIIKTGGKQVKVEVGQAIYVEKLDVEAGAEVTFNEVVLVGGETTKVGTPVVEGATV 60
           MSTYAIIKTGGKQVKVEVGQAIYVEK+D EAGAEVTFNEVVLVGG+ T VGTPVVEGATV
Sbjct: 1   MSTYAIIKTGGKQVKVEVGQAIYVEKIDAEAGAEVTFNEVVLVGGDKTVVGTPVVEGATV 60

Query: 61  VGTVEKQGKQKKVVSYKYKPKKGSHRKQGHRQPYTKVVINAINA 104
           VGTVEKQGKQKKVV++KYKPKKGSHRKQGHRQPYTKVVINAINA
Sbjct: 61  VGTVEKQGKQKKVVTFKYKPKKGSHRKQGHRQPYTKVVINAINA 104
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 848

A DNA sequence (GBSx0900) was identified in *S.agalactiae* <SEQ ID 2565> which encodes the amino acid sequence <SEQ ID 2566>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1032(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9369> which encodes amino acid sequence <SEQ ID 9370> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14809 GB:Z99118 excinuclease ABC (subunit C) [Bacillus subtilis]
 Identities = 221/373 (59%), Positives = 288/373 (76%)

Query: 1   MKSAAMTMEFERAAEYRDLIEAISLLRTKQRVIHQDMKDRDVFGYFVDKGWMCVQVFFVR 60
           M  AA  +EFERA E RD I  I     KQ++   D+ DRDVF Y  DKGWMCVQVFF+R
Sbjct: 206 MHEAAENLEFERAKELRDQIAHIESTMEKQKMTMNDLVDRDVFAYAYDKGWMCVQVFFIR 265

Query: 61  NGKLIQRDVNMFPYYNEPEEDFLTYIGQFYQDTKHFLPKEVFIPQDIDAKSVETIVGCKI 120
            GKLI+RDV+MFP Y E +E+FLT+IGQFY    HFLPKE+ +P  ID   +E ++   +
Sbjct: 266 QGKLIERDVSMFPLYQEADEEFLTFIGQFYSKNNHFLPKEILVPDSIDQSMIEQLLETNV 325

Query: 121 VKPQRGEKKQLVNLAIKNARVSLQQKFDLLEKDIRKTHGAIENLGNLLNIPKPVRIEAFD 180
            +P++G KK+L+ LA KNA+++L++KF L+E+D  ++ GA++ LG  LNI  P RI AFD
Sbjct: 326 HQPKKGPKKELLMLAHKNAKIALKEKFSLIERDEERSIGAVQKLGEALNIYTPHRIVAFD 385

Query: 181 NSNIQGTSPVAAMVVFVNGKPSKKDYRKFKIKTVIGPDDYASMREVIHRRYSRVLKDGLT 240
           NSNIQGT+PV+AM+VF++GKP KK+YRK+KIKTV GPDDY SMREV+ RRY+RVL++ L
Sbjct: 386 NSNIQGTNPVSAMIVFIDGKPYKKEYRKYKIKTVTGPDDYGSMREVVRRRYTRVLRENLP 445

Query: 241 PPDLIVIDGGQGQVNIARDVIENQFGLAIPIAGLQKNDKHQTHELLFGDPLEVVELPRNS 300
            PDLI+IDGG+GQ+N ARDVIEN+ GL IPIAGL K++KH+T  LL GDPLEV  L RNS
Sbjct: 446 LPDLIIIDGGKGQINAARDVIENELGLDIPIAGLAKDEKHRTSNLLIGDPLEVAYLERNS 505

Query: 301 EEFFLLHRIQDEVHRFAITFHRQLRSKNSFSSKLDGITGLGPKRKQLLMKHFKSLPNIQK 360
```

```
           +EF+LL RIQDEVHRFAI+FHRQ+R K++F S LD I G+G KRK++L+KHF S+  +++
Sbjct: 506 QEFYLLQRIQDEVHRFAISFHRQIRGKSAFQSVLDDIPGIGEKRKKMLLKHFGSVKKMKE 565

Query: 361 AEIEDIIMCGIPR 373
           A +EDI   G+P+
Sbjct: 566 ASLEDIKKAGVPQ 578
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2567> which encodes the amino acid sequence <SEQ ID 2568>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4332(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 289/385 (75%), Positives = 334/385 (86%)

Query: 1   MKSAAMTMEFERAAEYRDLIEAISLLRTKQRVIHQDMKDRDVFGYFVDKGWMCVQVFFVR 60
           M +A+  M FERAAEYRDLI  I+ +RTKQRV+ +D++DRD+FGY+VDKGWMCVQVFFVR
Sbjct: 206 MLAASKEMAFERAAEYRDLISGIATMRTKQRVMSKDLQDRDIFGYYVDKGWMCVQVFFVR 265

Query: 61  NGKLIQRDVNMFPYYNEPEEDFLTYIGQFYQDTKHFLPKEVFIPQDIDAKSVETIVGCKI 120
            GKLIQRDVN+FPYY + EEDFLTY+GQFYQD +HF+PKEVFIP+ ID + V  IV  KI
Sbjct: 266 QGKLIQRDVNLFPYYTDAEEDFLTYMGQFYQDKQHFIPKEVFIPEAIDEELVAAIVPTKI 325

Query: 121 VKPQRGEKKQLVNLAIKNARVSLQQKFDLLEKDIRKTHGAIENLGNLLNIPKPVRIEAFD 180
           +KP+RGEKKQLV LA KNARVSLQQKFDLLEKDI+KT GAIENLG LL I KPVRIEAFD
Sbjct: 326 IKPKRGEKKQLVALATKNARVSLQQKFDLLEKDIKKTSGAIENLGQLLRIDKPVRIEAFD 385

Query: 181 NSNIQGTSPVAAMVVFVNGKPSKKDYRKFKIKTVIGPDDYASMREVIHRRYSRVLKDGLT 240
           NSNIQGTSPVAAMVVFV+GKPSKKDYRKFKIKTV+GPDDYASMREV+ RRYSRV K+GL
Sbjct: 386 NSNIQGTSPVAAMVVFVDGKPSKKDYRKFKIKTVVGPDDYASMREVLFRRYSRVKKEGLQ 445

Query: 241 PPDLIVIDGGQGQVNIARDVIENQFGLAIPIAGLQKNDKHQTHELLFGDPLEVVELPRNS 300
            P+LI++DGG GQVN+A+DVIE Q GL IP+AGLQKNDKHQTH+LLFG+PLEVV LPR S
Sbjct: 446 APNLIIVDGGVGQVNVAKDVIEKQLGLTIPVAGLQKNDKHQTHDLLFGNPLEVVPLPRRS 505

Query: 301 EEFFLLHRIQDEVHRFAITFHRQLRSKNSFSSKLDGITGLGPKRKQLLMKHFKSLPNIQK 360
           EEFFLLHRIQDEVHRFA+TFHRQ+R KNSFSS LD I+GLGPKRKQLL++HFK++  I
Sbjct: 506 EEFFLLHRIQDEVHRFAVTFHRQVRRKNSFSSTLDHISGLGPKRKQLLLRHFKTITAIAS 565

Query: 361 AEIEDIIMCGIPRTVAESLRDSLND 385
           A  E+I   GIP+TV E+++   + D
Sbjct: 566 ATSEEIQALGIPKTVVEAIQQQITD 590
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 849

A DNA sequence (GBSx0901) was identified in *S.agalactiae* <SEQ ID 2569> which encodes the amino acid sequence <SEQ ID 2570>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2491(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 850

A DNA sequence (GBSx0902) was identified in *S.agalactiae* <SEQ ID 2571> which encodes the amino acid sequence <SEQ ID 2572>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3349(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA86651 GB:AB033763 glycerophosphoryl diester phosphodiesterase
           homologue [Staphylococcus aureus]
 Identities = 50/202 (24%), Positives = 96/202 (46%), Gaps = 15/202 (7%)

Query: 1   MDVIMTKDHKLVVIHDDNLKRLSGMNKDVSKLTLDQVTKIPIHQ---GRFA-SHIPSFTE 56
           +DV +TKD +L++IHDD L+R + M+ ++++L  D++           +F   H+P+F +
Sbjct: 36  LDVAITKDEQLIIIHDDYLERTTNMSGEITELNYDEIKDASAGSWFGEKFKDEHLPTFDD 95

Query: 57  FMKTAQSLDQKIMIELKPY-NQNLDIYADEFIKEFKE----LRLSTKHKVMSLNLTLIEK 111
            +K A   +  + +ELK     N    +   +K+ +E    L  + +  + S N+ L++
Sbjct: 96  VVKIANEYNMNLNVELKGITGPNGLALSKSMVKQVEEQLTNLNQNQEVLISSFNVVLVKL 155

Query: 112 VEKKLPQLDTGYLIPL-----HWGTLQNH-NVDFYGIEEFSYNDWIAYLAQEYNKQLYVW 165
            E+ +PQ +   +         W TL ++ N      E+         + +E   +L VW
Sbjct: 156 AEEIMPQYNRAVIFHTTSFREDWRTLLDYCNAKIVNTEDAKLTKAKVKMVKEAGYELNVW 215

Query: 166 TINRDNLMIRYLQSPVNGIITD 187
           T+N+     +     V+GI TD
Sbjct: 216 TVNKPARANQLANWGVDGIFTD 237
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2573> which encodes the amino acid sequence <SEQ ID 2574>. Analysis of this protein sequence reveals the following:

```
     Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.26   Transmembrane  239 - 255 ( 227 - 260)
   INTEGRAL    Likelihood = -9.45   Transmembrane   80 -  96 (  78 - 108)
   INTEGRAL    Likelihood = -9.13   Transmembrane  137 - 153 ( 131 - 160)
   INTEGRAL    Likelihood = -4.94   Transmembrane  278 - 294 ( 277 - 295)
   INTEGRAL    Likelihood = -3.56   Transmembrane   36 -  52 (  33 -  55)
   INTEGRAL    Likelihood = -3.56   Transmembrane  188 - 204 ( 185 - 206)
   INTEGRAL    Likelihood = -3.35   Transmembrane  314 - 330 ( 310 - 331)

----- Final Results -----
               bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB12801 GB:Z99109 similar to glycerophosphodiester
```

```
           phosphodiesterase [Bacillus subtilis]
 Identities = 67/244 (27%), Positives = 110/244 (44%), Gaps = 14/244 (5%)

Query: 344 VIAHRGLVSAGVENSLEALEGAKKAGSDYVELDLILTKDNHFVVSHDNRLKRLAGVNKTI 403
           +IAHRG      EN++ A + A K  +D +ELD+ LTKD   VV HD+R+ R    +  +
Sbjct: 3   IIAHRGASGYAPENTIAAFDLAVKMNADMIELDVQLTKDRQIVVIHDDRVDRTTNGSGFV 62

Query: 404 RNLTLKEVEHLTSHQGH---FSGRFVSFDTFYQKAKKLNMPLLIELKPIGTEPGNYVDLF 460
           ++ TL+E++ L +   +   F G  +       K     + LLIELK   ++ G   ++
Sbjct: 63  KDFTLEELQKLDAGSWYGPAFQGERIPTLEAVLKRYHKKIGLLIELKGHPSQVGIEEEVG 122

Query: 461 LETYHRLGISKDNKVMSLDLEVIEAIKKKNPSITTGYIIPIQFGFFG-------DEFVDF 513
            +   +   S +N V S     ++  ++  PSI T  I    FG            F ++
Sbjct: 123 -QLLGQFSFSINNIVQSFQFRSVQRFRELYPSIPTAVITRPNFGMLSRNQMKAFRSFANY 181

Query: 514 YVIEDFSYRSYLSSQAFWNNKEIYVWTINDPKRIEHYLLKPIQGIITDQPALTNQLIKDL 573
             I+       +      N   I+ WT+N+ K         + GI+TD P   + +IKD
Sbjct: 182 VNIKHTRLNRLMIGSINKNGLNIFAWTVNNQKTAAKLQAMGVDGIVTDYP---DFIIKDG 238

Query: 574 KQDN 577
           K +N
Sbjct: 239 KHEN 242
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 90/215 (41%), Positives = 136/215 (62%)

Query: 1   MDVIMTKDHKLVVIHDDNLKRLSGMNKDVSKLTLDQVTKIPIHQGRFASHIPSFTEFMKT 60
           +D+I+TKD+  VV HD+ LKRL+G+NK +  LTL +V  +  HQG F+    SF  F +
Sbjct: 375 LDLILTKDNHFVVSHDNRLKRLAGVNKTIRNLTLKEVEHLTSHQGHFSGRFVSFDTFYQK 434

Query: 61  AQSLDQKIMIELKPYNQNLDIYADEFIKEFKELRLSTKHKVMSLNLTLIEKVEKKLPQLD 120
           A+ L+  ++IELKP       Y D F++ +  L +S  +KVMSL+L +IE ++KK P +
Sbjct: 435 AKKLNMPLLIELKPIGTEPGNYVDLFLETYHRLGISKDNKVMSLDLEVIEAIKKKNPSIT 494

Query: 121 TGYLIPLHWGTLQNHNVDFYGIEEFSYNDWIAYLAQEYNKQLYVWTINRDNLMIRYLQSP 180
           TGY+IP+ +G   +  VDFY IE+FSY  +++  A   NK++YVWTIN    +  YL  P
Sbjct: 495 TGYIIPIQFGFFGDEFVDFYVIEDFSYRSYLSSQAFWNNKEIYVWTINDPKRIEHYLLKP 554

Query: 181 VNGIITDELNLFKVINKDIKNSPNYYQRALQLIDS 215
           + GIITD+  L   + KD+K   +Y+ R +++I S
Sbjct: 555 IQGIITDQPALTNQLIKDLKQDNSYFSRLVRIISS 589
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 851

A DNA sequence (GBSx0903) was identified in *S.agalactiae* <SEQ ID 2575> which encodes the amino acid sequence <SEQ ID 2576>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-15.02    Transmembrane    84 - 100 (  76 - 112)
   INTEGRAL    Likelihood = -3.50    Transmembrane   139 - 155 ( 139 - 157)
   INTEGRAL    Likelihood = -2.23    Transmembrane    41 -  57 (  39 -  59)
   INTEGRAL    Likelihood = -0.96    Transmembrane   179 - 195 ( 179 - 195)

----- Final Results -----
               bacterial membrane --- Certainty=0.7007(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9901> which encodes amino acid sequence <SEQ ID 9902> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 2574.

A related GBS gene <SEQ ID 8671> and protein <SEQ ID 8672> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      -3.38
GvH: Signal Score (-7.5): -4.08
     Possible site: 53
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 4 value: -15.02 threshold:  0.0
   INTEGRAL    Likelihood =-15.02   Transmembrane   84 - 100 (  76 - 112)
   INTEGRAL    Likelihood = -3.50   Transmembrane  139 - 155 ( 139 - 157)
   INTEGRAL    Likelihood = -2.23   Transmembrane   41 -  57 (  39 -  59)
   INTEGRAL    Likelihood = -0.96   Transmembrane  179 - 195 ( 179 - 195)
   PERIPHERAL  Likelihood =  2.01     104
 modified ALOM score:   3.50

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.7007(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 852

A DNA sequence (GBSx0904) was identified in *S.agalactiae* <SEQ ID 2577> which encodes the amino acid sequence <SEQ ID 2578>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4150(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 853

A DNA sequence (GBSx0905) was identified in *S.agalactiae* <SEQ ID 2579> which encodes the amino acid sequence <SEQ ID 2580>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane    2 -  18 (   2 -  18)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 854

A DNA sequence (GBSx0906) was identified in *S.agalactiae* <SEQ ID 2581> which encodes the amino acid sequence <SEQ ID 2582>. This protein is predicted to be nad(p)h nitroreductase ydgi. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.81    Transmembrane  127 - 143 ( 126 - 143)

----- Final Results -----
              bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAC09964 GB:AX033132 unnamed protein product [Bacillus subtilis]
 Identities = 62/204 (30%), Positives = 106/204 (51%), Gaps = 11/204 (5%)

Query: 3   FLELNKKRHAVKHFNDKPVDFKDVRTAI-EIATLAPSANNIQPWKFVVVQ--EKKSALAE 59
           F+E+ K R ++++++      K+  T I E AT APS+ N QPW+F+V+   E K  LA
Sbjct: 7   FMEIMKGRRSIRNYDPAVKISKEEMTEILEEATTAPSSVNAQPWRFLVIDSPEGKEKLAP 66

Query: 60  GLPESNCNQINQAQYVIALFTDTD----LGQRSRKIARIGRRSLPDDLIGYYMETLPPRY 115
            L   N  Q+  +  VIA+F D +    L +   K   +G   +P ++    +  L   +
Sbjct: 67  -LASFNQTQVTTSSAVIAVFADMNNADYLEEIYSKAVELG--YMPQEVKDRQIAALTAHF 123

Query: 116 ALYSEKQTGEYLSLNAGIVAMNLVLALTDQGISSNMILGFDKAITNDVLEIDK-RFRPEI 174
                +   E + ++ G+V+M L+L     G  +N I G+DK    +   +DK R+ P +
Sbjct: 124 EKLPAQVNRETILIDGGLVSMQLMLTARAHGYDTNPIGGYDKENIAETFGLDKERYVPVM 183

Query: 175 LITVGYSDEKVEPSYRLPVDHIIE 198
           L+++G + ++   SYRLP+D I E
Sbjct: 184 LLSIGKAADEGYASYRLPIDTIAE 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2583> which encodes the amino acid sequence <SEQ ID 2584>. Analysis of this protein sequence reveals the following:

```
     Possible site: 38

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.18    Transmembrane  127 - 143 ( 126 - 143)

----- Final Results -----
              bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAC09964 GB:AX033132 unnamed protein product [Bacillus subtilis]
 Identities = 63/204 (30%), Positives = 109/204 (52%), Gaps = 11/204 (5%)

Query: 3   FLELNKKRHAIKTFNDQ-PVDYEDLRTAIEIATLAPSANNIQPWKFVVVQ--EKKAELAK 59
           F+E+ K R +I+ ++    +  E++   +E AT APS+ N QPW+F+V+   E K +LA
```

```
Sbjct: 7   FMEIMKGRRSIRNYDPAVKISKEEMTEILEEATTAPSSVNAQPWRFLVIDSPEGKEKLA- 65

Query: 60  GLPLA--NKVQVEQAQYVVALFSDTDLALRSRKIARIGVK--SLPDDLIGYYMETLPPRF 115
              PLA  N+ QV  +  V+A+F+D + A    +I    V+   +P ++    +  L   F
Sbjct: 66  --PLASFNQTQVTTSSAVIAVFADMNNADYLEEIYSKAVELGYMPQEVKDRQIAALTAHF 123

Query: 116 AAFNEVQTGEYLAINAGIVAMNLVLSLTDQKIASNIILGFDKSTTNEILDID-PRFRPEL 174
                     E + I+ G+V+M L+L+       +N I G+DK    E   +D  R+ P +
Sbjct: 124 EKLPAQVNRETILIDGGLVSMQLMLTARAHGYDTNPIGGYDKENIAETFGLDKERYVPVM 183

Query: 175 LITVGYSDEKPEPSYRLPVDEVIE 198
            L+++G + ++   SYRLP+D + E
Sbjct: 184 LLSIGKAADEGYASYRLPIDTIAE 207
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 157/200 (78%), Positives = 184/200 (91%)

Query: 1   MKFLELNKKRHAVKHFNDKPVDFKDVRTAIEIATLAPSANNIQPWKFVVVQEKKSALAEG 60
            MKFLELNKKRHA+K FND+PVD++D+RTAIEIATLAPSANNIQPWKFVVVQEKK+ LA+G
Sbjct: 1   MKFLELNKKRHAIKTFNDQPVDYEDLRTAIEIATLAPSANNIQPWKFVVVQEKKAELAKG 60

Query: 61  LPESNCNQINQAQYVIALFTDTDLGQRSRKIARIGRRSLPDDLIGYYMETLPPRYALYSE 120
            LP +N  Q+ QAQYV+ALF+DTDL  RSRKIARIG +SLPDDLIGYYMETLPPR+A ++E
Sbjct: 61  LPLANKVQVEQAQYVVALFSDTDLALRSRKIARIGVKSLPDDLIGYYMETLPPRFAAFNE 120

Query: 121 KQTGEYLSLNAGIVAMNLVLALTDQGISSNMILGFDKAITNDVLEIDKRFRPEILITVGY 180
             QTGEYL++NAGIVAMNLVL+LTDQ I+SN+ILGFDK+ TN++L+ID RFRPE+LITVGY
Sbjct: 121 VQTGEYLAINAGIVAMNLVLSLTDQKIASNIILGFDKSTTNEILDIDPRFRPELLITVGY 180

Query: 181 SDEKVEPSYRLPVDHIIEKR 200
            SDEK EPSYRLPVD +IE+R
Sbjct: 181 SDEKPEPSYRLPVDEVIERR 200
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 855

A DNA sequence (GBSx0907) was identified in *S.agalactiae* <SEQ ID 2585> which encodes the amino acid sequence <SEQ ID 2586>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2895(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45369 GB:U78036 dipeptidase [Lactococcus lactis]
 Identities = 312/474 (65%), Positives = 370/474 (77%), Gaps = 11/474 (2%)

Query: 2   TIDFRAEVDKRKDALMDDLINLLRINSERDDSQADAEHPFGPGPVKALEFFLEMAERDGY 61
            TIDF+AEV+KRKDALM+DL +LLRI+S  D   ADAE+PFGPGP KAL+ FL++AERDGY
Sbjct: 3   TIDFKAEVEKRKDALMEDLFSLLRIDSAMDMEHADAENPFGPGPRKALDAFLKIAERDGY 62

Query: 62  ETKNVDNYAGHFTFGQGE----EELGIFGHLDVVPAGSGWDTDPYEPVIKDNRLYARGSS 117
             TKN DNY GHF +  G     E LGI GHLDVVPAGSGWD++P+EP I++  LYARG+S
Sbjct: 63  TTKNYDNYVGHFEYENGANADAEVLGIIGHLDVVPAGSGWDSNPFEPEIRNGNLYARGAS 122

Query: 118 DDKGPTMACYYALKIIKELGLPTSKKVRFVVGTDEESGWGDMDYYFEHVGLPKPDFGFSP 177
            DDKGPT+ACYYALKI+KEL LP SKK+RF+VGT+EE+GW DMDYYFEH  LP PDFGFSP
Sbjct: 123 DDKGPTVACYYALKILKELNLPLSKKIRFIVGTNEETGWADMDYYFEHCELPLPDFGFSP 182
```

```
Query: 178 DAEFPIINGEKGNITEYLHFSGENKGAVRLHSFSGGLRENMVPESATARFTSHLDQTTLG 237
           DAEFPIINGEKGNITEYLHFSG+N G V LHSF  GL ENMVPESATA  +   D   L
Sbjct: 183 DAEFPIINGEKGNITEYLHFSGKNAGQVVLHSFKAGLAENMVPESATAVISGAKD---LE 239

Query: 238 ASLADFASKH---NLKAELSVEDEQYTATVYGKSAHGSTPQEGVNGATYLALYLSQFDFE 294
           A+L  F ++H   NL+ +L   D + T T+YGKSAHG+ P++G+NGATYL L+L+QFDF
Sbjct: 240 AALEKFVAEHASKNLRFDLEEADGKATITLYGKSAHGAMPEKGINGATYLTLFLNQFDFA 299

Query: 295 GPARAFLDVTANIIHEDFSGEKLGVAYEDDCMGPLSMNAGVFQFDETNDDNTIALNFRYP 354
             A AF+ V A  + ED  GEKLG A+ D+ M   SMNAGV+ FDE N +  IALNFR+P
Sbjct: 300 DGAAAFIKVGAEKLLEDHEGEKLGTAFVDELMENTSMNAGVWSFDE-NGEGKIALNFRFP 358

Query: 355 QGTDAKTIQTKLEKLNGVEKVTLSDHEHTPHYVPMDDELVSTLLAVYEKQTGLKGHEQVI 414
           QG   + +Q  L KL+GV +V LS H HTPHYVPM D LVSTL+ VYEK TGLKG+E +I
Sbjct: 359 QGNSPERMQEILAKLDGVVEVELSKHLHTPHYVPMSDPLVSTLIDVYEKHTGLKGYETII 418

Query: 415 GGGTFGRLLERGVAYGAMFPGDENTMHQANEYMPLENIFRSAAIYAEAIYELIK 468
           GGGTFGRLLERGVAYGAMF G+ ++MHQANE  P+ENI+++A IYAEAIYEL K
Sbjct: 419 GGGTFGRLLERGVAYGAMFEGEPDSMHQANEMKPVENIYKAAVIYAEAIYELAK 472
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2587> which encodes the amino acid sequence <SEQ ID 2588>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3107(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 361/467 (77%), Positives = 403/467 (85%)

Query: 2   TIDFRAEVDKRKDALMDDLINLLRINSERDDSQADAEHPFGPGPVKALEFFLEMAERDGY 61
           TIDF+AEVDKRK A++ DL++LLRINSERDD  AD +HPFGPGPVKALE FL MAERDGY
Sbjct: 20  TIDFKAEVDKRKKAMLADLVDLLRINSERDDQLADDKHPFGPGPVKALEHFLAMAERDGY 79

Query: 62  ETKNVDNYAGHFTFGQGEEELGIFGHLDVVPAGSGWDTDPYEPVIKDNRLYARGSSDDKG 121
           +T+N+DNYAG F FGQG+E LGIFGHLDVVPAGSGWDTDPYEPVIKD+R+YARGSSDDKG
Sbjct: 80  KTRNIDNYAGDFEFGQGDEVLGIFGHLDVVPAGSGWDTDPYEPVIKDDRIYARGSSDDKG 139

Query: 122 PTMACYYALKIIKELGLPTSKKVRFVVGTDEESGWGDMDYYFEHVGLPKPDFGFSPDAEF 181
           PTMACYYALKIIKELGLP SKKVRF+VGTDEESGWGDMDYYF H GL  PDFGFSPDAEF
Sbjct: 140 PTMACYYALKIIKELGLPVSKKVRFIVGTDEESGWGDMDYYFAHNGLKNPDFGFSPDAEF 199

Query: 182 PIINGEKGNITEYLHFSGENKGAVRLHSFSGGLRENMVPESATARFTSHLDQTTLGASLA 241
           PIINGEKGNITEYLHF+G+NKGA  LH F GGLRENMVPESATA  T+  D   L A+L
Sbjct: 200 PIINGEKGNITEYLHFAGDNKGAFVLHRFQGGLRENMVPESATAVITAPHDLDVLEAALE 259

Query: 242 DFASKHNLKAELSVEDEQYTATVYGKSAHGSTPQEGVNGATYLALYLSQFDFEGPARAFL 301
            F S+H +K  +   D +   T+ GKSAHGSTP+ GVNGAT LA +L+QF FEG A+ +L
Sbjct: 260 QFLSEHGVKGSMKATDGKIEVTIIGKSAHGSTPEAGVNGATLLAKFLNQFTFEGAAKDYL 319

Query: 302 DVTANIIHEDFSGEKLGVAYEDDCMGPLSMNAGVFQFDETNDDNTIALNFRYPQGTDAKT 361
            V   ++HEDF+ EKLG+AY DD MG LSMNAGVF FD  + DNTIALNFRYP+GTDA T
Sbjct: 320 HVAGEVLHEDFAAEKLGLAYTDDRMGALSMNAGVFTFDSQSADNTIALNFRYPKGTDAAT 379

Query: 362 IQTKLEKLNGVEKVTLSDHEHTPHYVPMDDELVSTLLAVYEKQTGLKGHEQVIGGGTFGR 421
           ++  LEKL G+ KV+LS+HEHTPHYVPMDDELV+TLLAVYEKQTGLKG+EQVIGGGTFGR
Sbjct: 380 LKAGLEKLPGLTKVSLSEHEHTPHYVPMDDELVATLLAVYEKQTGLKGYEQVIGGGTFGR 439

Query: 422 LLERGVAYGAMFPGDENTMHQANEYMPLENIFRSAAIYAEAIYELIK 468
           LLERGVA+GAMFPGDENTMHQANEYMPLENI+RSAAIYAEAIYELIK
Sbjct: 440 LLERGVAFGAMFPGDENTMHQANEYMPLENIYRSAAIYAEAIYELIK 486
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 856

A DNA sequence (GBSx0908) was identified in *S.agalactiae* <SEQ ID 2589> which encodes the amino acid sequence <SEQ ID 2590>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5598(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC21888 GB:U32707 H. influenzae predicted coding region
           HI0220.2 [Haemophilus influenzae Rd]
 Identities = 123/192 (64%), Positives = 160/192 (83%), Gaps = 1/192 (0%)

Query: 1   MTDLEKIIKAIKSDSQNQNYTENGIDPLFAAPKTARINIVGQAPGLKTQEARLYWKDKSG 60
           + +L++I  +I +D QN+++TE GI PLF+APKTARINIVGQAPGLK +++RLYW DKSG
Sbjct: 21  LKNLDEITSSIIADPQNKDFTERGIFPLFSAPKTARINIVGQAPGLKAEQSRLYWNDKSG 80

Query: 61  DRLRQWLGVDEETFYHSGKFAVLPLDFYYPGKGKSGDLSPRKGFAEKWHPLILKEMPNVQ 120
           DRLR+WLGVD + FY+SG FAVLP+DFYYPG GKSGDL PR+GFAE+WHP+IL  +PN+Q
Sbjct: 81  DRLREWLGVDYDYFYNSGIFAVLPMDFYYPGYGKSGDLPPRQGFAERWHPMILGNLPNIQ 140

Query: 121 LTLLVGQYTQKYYLGSSAHKNLTETVKAYKDYLPDYLPLVHPSPRNQIWLKKNPWFEKDL 180
           LT+L+GQY QKYYL  +   N+T TVK Y+ +LP ++PLVHPSPRNQ+W+ KNPWFE+ +
Sbjct: 141 LTILIGQYAQKYYLPEN-KDNVTNTVKNYRQFLPHFMPLVHPSPRNQLWVTKNPWFEEQV 199

Query: 181 IVDLQKIVADIL 192
           I +LQ +V  I+
Sbjct: 200 IPELQILVKQII 211
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2591> which encodes the amino acid sequence <SEQ ID 2592>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3740(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 122/189 (64%), Positives = 150/189 (78%)

Query: 4   LEKIIKAIKSDSQNQNYTENGIDPLFAAPKTARINIVGQAPGLKTQEARLYWKDKSGDRL 63
           ++ + KAI +D  N +YTE GI PL+ AP+TARI IVGQAPG+  Q  +LYW D+SG RL
Sbjct: 1   MDDLTKAIMADEANLSYTERGIFPLYDAPQTARIIIVGQAPGIVAQGTKLYWNDRSGIRL 60

Query: 64  RQWLGVDEETFYHSGKFAVLPLDFYYPGKGKSGDLSPRKGFAEKWHPLILKEMPNVQLTL 123
           R WLGVD +TFYHSG F ++P+DFYYPGKGKSGDL PR+GFA KWHP +   MP V+LT+
Sbjct: 61  RDWLGVDNDTFYHSGLFGIIPMDFYYPGKGKSGDLPPREGFAAKWHPPLRALMPEVELTI 120

Query: 124 LVGQYTQKYYLGSSAHKNLTETVKAYKDYLPDYLPLVHPSPRNQIWLKKNPWFEKDLIVD 183
           LVG+Y Q +YLG+ A+K LTETV+ ++DYLPDY PLVHPSPRNQ+WL KNPWFE+DL+
Sbjct: 121 LVGRYAQDFYLGNKAYKTLTETVRHFEDYLPDYFPLVHPSPRNQLWLAKNPWFEQDLLPI 180
```

```
Query: 184 LQKIVADIL 192
           LQK V  IL
Sbjct: 181 LQKRVEAIL 189
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 857

A DNA sequence (GBSx0909) was identified in *S.agalactiae* <SEQ ID 2593> which encodes the amino acid sequence <SEQ ID 2594>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4178(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 858

A DNA sequence (GBSx0910) was identified in *S.agalactiae* <SEQ ID 2595> which encodes the amino acid sequence <SEQ ID 2596>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2779(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9899> which encodes amino acid sequence <SEQ ID 9900> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD35886 GB:AE001748 conserved hypothetical protein [Thermotoga maritima]
 Identities = 36/124 (29%), Positives = 58/124 (46%), Gaps = 3/124 (2%)

Query: 19  VPTKELLADYFNRMEFAIGRVEAHVLAHFDYGFRKLNLDVEDLKPFETQLKRIFIKMLSK 78
            +P  EL  DY  R  F + RV+ H LAH DY  R    D    K     +++I + ++
Sbjct: 98  LPPDELARDYLERTLFVMERVKFHTLAHLDYPARYAKAD---FKANRDLIEKILVFLVKN 154

Query: 79  GLAFELNTKSLYLYGNEKLYRYALEILKQLGCKQYSIGSDGHIPEHFCYEFDRLQGLLKD 138
              A E+NT  L+ +G      + +E+   LG +  +IGSD H  +H     + +   LK
Sbjct: 155 EKALEINTAGLFKHGKPNPDYWIVEMYYDLGGRVVTIGSDAHESQHIGRGIEEVMRELKK 214

Query: 139 YQID 142
            +  +
Sbjct: 215 FNFE 218
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 859

A DNA sequence (GBSx0911) was identified in *S.agalactiae* <SEQ ID 2597> which encodes the amino acid sequence <SEQ ID 2598>. This protein is predicted to be alkaline amylopullulanase (pulA). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.08   Transmembrane 1225 -1241 (1222 -1247)
   INTEGRAL    Likelihood = -2.44   Transmembrane   19 -  35 (  18 -  36)
   INTEGRAL    Likelihood = -0.11   Transmembrane 1146 -1162 (1146 -1162)

----- Final Results -----
               bacterial membrane --- Certainty=0.5034(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG33958 GB:AF217414 pullulanase [Streptococcus pneumoniae]
 Identities = 641/1311 (48%), Positives = 854/1311 (64%), Gaps = 88/1311 (6%)

Query: 1    MKRKDLFGDKQTQYTIRKLSVGVASVATGVCIFLHSPQVFAEEVSASPANTAIAESNINQ 60
            M++     +K+  Y+IR L  G  SV  G  + L                 A+A   I+
Sbjct: 1    MRKTPSHTEKKMVYSIRSLKNGTGSVLIGASLVL----------------LAMATPTISS 44

Query: 61   VDNQQSTNLKDDINSNSETVVTPSDMPDTKQLVSDETDTQKGVTEPDKATSLLEENKG-P 119
             ++  +TN  +  N N+ T+  P  + DT    +     +  ++ P  A + LE+ +  P
Sbjct: 45   DESTPTTN--EPNNRNTTTLAQP--LTDT---AAGSGKNESDISSPGNANASLEKTEEKP 97

Query: 120  VSDKNTLDLKVAPSTLQNTPDKTSQAIGAPSPTLKVANQAPRIENGYFRLHLKELPQGHP 179
             ++  T     A    Q   D++S+   + SP          IE+ YFR+H+K+LP+ +
Sbjct: 98   ATEPTTPAASPADPAPQTGQDRSSEPTTSTSPVTTETKAEEPIEDNYFRIHVKKLPEENK 157

Query: 180  VESTGLWIWGDVDQPSSNWPNGAIPMTDAKKDDYGYYVDFKLSEKQRKQISFLINNKAGT 239
             ++ GLW W DV++PS NWPNGA+   DAKKDDYGYY+D KL  +Q K+ISFLINN AG
Sbjct: 158  -DAQGLWTWDDVEKPSENWPNGALSFKDAKKDDYGYYLDVKLKGEQAKKISFLINNTAGK 216

Query: 240  NLSGDHHIPLLRPEMNQVWIDEKYGTHTYQPLKEGYVRINYLSSSSNYDHLSAWLFKDVA 299
            NL+GD  +  L P+MN+ W+D+ Y   +Y+P   G VR+NY  +  NYD  S W + DV
Sbjct: 217  NLTGDKSVEKLVPKMNEAWLDQDYKVFSYEPQPAGTVRVNYYRTDGNYDKKSLWYWGDVK 276

Query: 300  TPSTT-WPDGSNFVNQGLYGRYIDVSLKTNAKEIGFLILDESKTGDAVKVQPNDYVFRDL 358
             PS+  WPDG++F   G YGRYID+ L   A+E GFL+LDESK GD VK++  +Y F DL
Sbjct: 277  NPSSAQWPDGTDFTATGKYGRYIDIPLNEAAREFGFLLLDESKQGDDVKIRKENYKFTDL 336

Query: 359  ANHNQIFVKDKDPKVYNNPYYIDQVQLKDAQQIDLTSIQASFTTLDGVDKTEILKELKVT 418
             NH+QIF+KD D  +Y NPYY+ +++   AQ +  +SI++SF+TL G  K +ILK   +T
Sbjct: 337  KNHSQIFLKDDDESIYTNPYYVHDIRMTGAQHVGTSSIESSFSTLVGAKKEDILKHSNIT 396

Query: 419  DKNQNAIQISDITLDTSKSLLIIKGDFNPKQGHFNISYNGNNVMTRQSWEFKDQLYAYSG 478
            +   N + I+D+ +D +   +   GDF+  +  + +SYN +   T+ SW  KD+ Y+Y G
Sbjct: 397  NHLGNKVTITDVAIDEAGKKVTYSGDFSDTKHPYTVSYNSDQFTTKTSWRLKDETYSYDG 456

Query: 479  NLGAVLNQDGSKVEASLWSPSADSVTMIIYDKDNQNRVVATTPLMKNNKGVWQTILDT-- 536
             LGA L ++G +V+ +LWSPSAD V++++YDK++ ++VV T  L K  +G W+  LD+
Sbjct: 457  KLGADLKEEGKQVDLTLWSPSADKVSVVVYDKNDPDKVVGTVALEKGERGTWKQTLDSTN 516

Query: 537  KLGIKNYTGYYYLYEIKRGKDKVKILDPYAKSLAEWDSNT--VNDDIKTAKAAFVNPSQL 594
            KLGI ++TGYYY Y+I+R    V  LDPYAKSLA W+S+   ++D  K AKAAFV+P++L
Sbjct: 517  KLGITDFTGYYYQYQIERQGKTVLALDPYAKSLAAWNSDDAKIDDAHKVAKAAFVDPAKL 576

Query: 595  GPQNLSFAKIANFKGRQDAVIYEAHVRDFTSDRSLDGKLKNQFGTFAAFSEKLDYLQKLG 654
```

```
            GPQ+L++ KI NFK R+DAVIYEAHVRDFTSD ++   L   FGTF AF EKLDYL+ LG
Sbjct: 577  GPQDLTYGKIHNFKTREDAVIYEAHVRDFTSDPAIAKDLTKPFGTFEAFIEKLDYLKDLG 636

Query: 655  VTHIQLLPVLSYFYVNEMDKSRSTA-YTSSDNNYNWGYDPQSYFALSGMYSEKPKDPSAR 713
            VTHIQLLPVLSY++VNE+      + Y SS++NYNWGYDPQ+YF+L+GMYS  PK+P  R
Sbjct: 637  VTHIQLLPVLSYYFVNELKNHEHLSDYASSNSNYNWGYDPQNYFSLTGMYSSDPKNPEKR 696

Query: 714  IAELKQLIHDIHKRGMGVILDVVYNHTAKTYLFEDIEPNYYHFMNEDGSPRESFGGGRLG 773
            IAE K LI++IHKRGMG ILDVVYNHTAK  +FED+EPNYYHFM+ DG+PR SFGGGRLG
Sbjct: 697  IAEFKNLINEIHKRGMGAILDVVYNHTAKVDIFEDLEPNYYHFMDADGTPRTSFGGGRLG 756

Query: 774  TTHAMSRRVLVDSIKYLTSEFKVDGFRFDMMGDHDAAAIELAYKEAKAINPNMIMIGEGW 833
            TTH M++R+LVDSIKYL   +KVDGFRFDMMGDHDAA+IE AYK A+A+NPN+IM+GEGW
Sbjct: 757  TTHHMTKRLLVDSIKYLVDTYKVDGFRFDMMGDHDAASIEEAYKAARALNPNLIMLGEGW 816

Query: 834  RTFQGDQGQPVKPADQDWMKSTDTVGVFSDDIRNSLKSGFPNEGTPAFITGGPQSLQGIF 893
            RT+ GD+  P K ADQDWMK TDTV VFSDDIRN+LKSG+PNEG PAFITGG + +  IF
Sbjct: 817  RTYAGDENMPTKAADQDWMKHTDTVAVFSDDIRNNLKSGYPNEGQPAFITGGKRDVNTIF 876

Query: 894  KNIKAQPGNFEADSPGDVVQYIAAHDNLTLHDVIAKSINKDPKVAEE--EIHRRLRLGNV 951
            KN+ AQP NFEADSPGDV+QYIAAHDNLTL D+IA+SI KDP  AE   EIHRRLRLGN+
Sbjct: 877  KNLIAQPTNFEADSPGDVIQYIAAHDNLTLFDIIAQSIKKDPSKAENYAEIHRRLRLGNL 936

Query: 952  MILTSQGTAFIHSGQEYGRTKRLLNPDYMTKVSDDKLPNKATLIEAVK----EYPYFIHD 1007
            M+LT+QGT FIHSGQEYGRTK+  NP Y T V++DK+PNK+ L+        +YPYFIHD
Sbjct: 937  MVLTAQGTPFIHSGQEYGRTKQFRNPAYRTPVAEDKVPNKSHLLRDKDGNPFDYPYFIHD 996

Query: 1008 SYDSSDAINHFDWAAATDNNKHPISTKTQAYTAGLITLRRSTDAFRKLSKAEIDREVSLI 1067
            SYDSSDA+N FDW  ATD   +P + K++ Y  GLI LR+STDAFR  S  +I   V LI
Sbjct: 997  SYDSSDAVNKFDWTKATDGKAYPENVKSRDYMKGLIALRQSTDAFRLKSLQDIKDRVHLI 1056

Query: 1068 TEVGQGDIKEKDLVIAYQTIDSKGDIYAVFVNADSKARNVLLGEKYKHLLKGQVIVDADQ 1127
            T  GQ  ++++D+VI YQ     GDIYAVFVNAD KAR   LG  + HL   +V+ D +Q
Sbjct: 1057 TVPGQNGVEKEDVVIGYQITAPNGDIYAVFVNADEKAREFNLGTAFAHLRNAEVLADENQ 1116

Query: 1128 AGIKPISTPRGVHFEKDSLLIDPLTAIVIKVGKVAPS---------------PKEELQAD 1172
            AG   I+ P+G+ + +  L ++ LTA V++V +   S               P+ + +A
Sbjct: 1117 AGSVGIANPKGLEWTEKGLKLNALTATVLRVSQNGTSHESTAEEKPDSTPSKPEHQNEAS 1176

Query: 1173 YPKTQ----------SFKESKTVEKVNRIANKT---------------SITPVVSKKADS 1207
            +P  Q          +  ++K  +  N+ +  T               S+   V K++
Sbjct: 1177 HPAHQDPAPEARPDSTKPDAKVADAENKPSQATADSQAEQPAQEAQASSVKEAVRKESVE 1236

Query: 1208 YLTNE---------ANLPKTGDKSSKILSVVGISILASLLALVGLSLKRNR 1249
              + E         A LP TG K+    L    GIS+LA LL L G  LK  +
Sbjct: 1237 NSSKENISATPDRQAELPNTGIKNENKLLFAGISLLA-LLGL-GFLLKNKK 1285
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2599> which encodes the amino acid sequence <SEQ ID 2600>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.83    Transmembrane 1153 -1169 (1148 -1171)
   INTEGRAL    Likelihood = -1.97    Transmembrane   29 -   45 (  28 -   46)

----- Final Results -----
               bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9125> which encodes the amino acid sequence <SEQ ID 9126>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 26
>>> Seems to have an uncleavable N-term signal seq
```

```
----- Final Results -----
              bacterial membrane --- Certainty= 0.533(Affirmative) < succ>
               bacterial outside --- Certainty= 0.000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

LPXTG motif: 1133-1137
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 715/1097 (65%), Positives = 872/1097 (79%), Gaps = 21/1097 (1%)

Query: 156   ANQAPRIENGYFRLHLKELPQGHPVESTGLWIWGDVDQPSSNWPNGAIPMTDAKKDDYGY 215
             AN A   E+ + R+H K LP G  + S GLW+WGDVDQPS +WPNGAI MT AKKDDYGY
Sbjct: 95    ANPASIAEH-HLRMHFKTLPAGESLGSLGLWVWGDVDQPSKDWPNGAITMTKAKKDDYGY 153

Query: 216   YVDFKLSEKQRKQISFLINNKAGTNLSGDHHIPLLRPEMNQVWIDEKYGTHTYQPLKEGY 275
             Y+D  L+ K R+Q+S+LINNKAG NLS D HI LL P+MN+VWIDE Y  H Y+PLK+GY
Sbjct: 154   YLDVPLAAKHRQQVSYLINNKAGENLSKDQHISLLTPKMNEVWIDENYHAHAYRPLKKGY 213

Query: 276   VRINYLSSSSNYDHLSAWLFKDVATPSTTWPDGSNFVNQGLYGRYIDVSLKTNAKEIGFL 335
             +RINY + S +YD+L+ W FKDV TP+T WP+G +  ++G YG Y+DV LK  A EIGFL
Sbjct: 214   LRINYHNQSGHYDNLAVWTFKDVKTPTTDWPNGLDLSHKGHYGAYVDVPLKEGANEIGFL 273

Query: 336   ILDESKTGDAVKVQPNDYVFRDLANHNQIFVKDKDPKVYNNPYYIDQVQLKDAQQIDLTS 395
             ILD+SKTGDA+KVQP DY+F++L NH Q+FVKD DPKVYNNPYYIDQV LK A+Q
Sbjct: 274   ILDKSKTGDAIKVQPKDYLFKELDNHTQVFVKDTDPKVYNNPYYIDQVSLKGAEQTTPNE 333

Query: 396   IQASFTTLDGVDKTEILKELKVTDKNQNAIQISDITLDTSKSLLIIKGDFNPKQGHFNIS 455
             I+A FTTLDG+D+  + + +K+TDK    + I ++TLD  KS++ +KGDF  +   + ++
Sbjct: 334   IKAIFTTLDGLDEDAVKQNIKITDKAGKTVAIDELTLDRDKSVMTLKGDFKAQGAVYTVT 393

Query: 456   YNGNNVMTRQSWEFKDQLYAYSGNLGAVLNQDGSKVEASLWSPSADSVTMIIYDKDNQNR 515
             +    + + RQSW+ KD+LYAY G LGA L +DGS V+ +LWSPSAD+V +++YDK +Q R
Sbjct: 394   FGEVSQVARQSWQLKDKLYAYDGELGATLAKDGS-VDLALWSPSADTVKVVVYDKQDQTR 452

Query: 516   VVATTPLMKNNKGVWQTIL--DTKLGIKNYTGYYYLYEIKRGKDKVKILDPYAKSLAEWD 573
             VV    L K++KGVW+  L  D+  GI +YTGYYYLYEI RG++KV +LDPYAKSLA W+
Sbjct: 453   VVGQADLTKSDKGVWRAHLTSDSVKGISDYTGYYYLYEITRGQEKVMVLDPYAKSLAAWN 512

Query: 574   SNTVNDDIKTAKAAFVNPSQLGPQNLSFAKIANFKGRQDAVIYEAHVRDFTSDRSLDGKL 633
               T  DDIKTAKAAF++PS+LGP  L FAKI NFK R+DA+IYEAHVRDFTSD++L+GKL
Sbjct: 513   DATATDDIKTAKAAFIDPSKLGPTGLDFAKINNFKKREDAIIYEAHVRDFTSDKALEGKL 572

Query: 634   KNQFGTFAAFSEKLDYLQKLGVTHIQLLPVLSYFYVNEMDKSRSTAYTSSDNNYNWGYDP 693
              + FGTF+AF E+LDYL+ LGVTH+QLLPVLSYFY NE+DKSRSTAYTSSDNNYNWGYDP
Sbjct: 573   THPFGTFSAFVEQLDYLKDLGVTHVQLLPVLSYFYANELDKSRSTAYTSSDNNYNWGYDP 632

Query: 694   QSYFALSGMYSEKPKDPSARIAELKQLIHDIHKRGMGVILDVVYNHTAKTYLFEDIEPNY 753
             Q YFALSGMYS  P DP+ RIAELK L+++IHKRGMGVI DVVYNHTA+TYLFED+EPNY
Sbjct: 633   QHYFALSGMYSANPNDPALRIAELKNLVNEIHKRGMGVIFDVVYNHTARTYLFEDLEPNY 692

Query: 754   YHFMNEDGSPRESFGGGRLGTTHAMSRRVLVDSIKYLTSEFKVDGFRFDMMGDHDAAAIE 813
             YHFMN DG+ RESFGGGRLGTTHAMSRR+LVDSI YLT EFKVDGFRFDMMGDHDAAAIE
Sbjct: 693   YHFMNADGTARESFGGGRLGTTHAMSRRILVDSITYLTREFKVDGFRFDMMGDHDAAAIE 752

Query: 814   LAYKEAKAINPNMIMIGEGWRTFQGDQGQPVKPADQDWMKSTDTVGVFSDDIRNSLKSGF 873
              A+K AKAINPN IMIGEGWRT+QGD+G+    ADQDWMK+T+TVGVFSDDIRN+LKSGF
Sbjct: 753   QAFKAAKAINPNTIMIGEGWRTYQGDEGKKEIAADQDWMKATNTVGVFSDDIRNTLKSGF 812

Query: 874   PNEGTPAFITGGPQSLQGIFKNIKAQPGNFEADSPGDVVQYIAAHDNLTLHDVIAKSINK 933
             PNEGT AFITGG ++L+G+FK IKAQPGNFEAD+PGDVVQYIAAHDNLTLHDVIAKSINK
Sbjct: 813   PNEGTAAFITGGAKNLEGLFKTIKAQPGNFEADAPGDVVQYIAAHDNLTLHDVIAKSINK 872

Query: 934   DPKVAEEEIHRRLRLGNVMILTSQGTAFIHSGQEYGRTKRLLNPDYMTKVSDDKLPNKAT 993
             DPKVAEEEIH+R+RLGN MILT+QGTAFIHSGQEYGRTK+LLNPDY TK SDDK+PNKAT
Sbjct: 873   DPKVAEEEIHKRIRLGNTMILTAQGTAFIHSGQEYGRTKQLLNPDYKTKASDDKVPNKAT 932

Query: 994   LIEAVKEYPYFIHDSYDSSDAINHFDWAAATDNNKHPISTKTQAYTAGLITLRRSTDAFR 1053
```

```
            LI+AV +YPYFIHDSYDSSDA+NHFDWA ATD+  HPIS +T+AYT GLI LRRSTDAF
Sbjct: 933  LIDAVAQYPYFIHDSYDSSDAVNHFDWAKATDSIAHPISNQTKAYTQGLIALRRSTDAFT 992

Query: 1054 KLSKAEIDREVSLITEVGQGDIKEKDLVIAYQTIDSKGDIYAVFVNADSKARNVLLGEKY 1113
            K +KAE+DR+V+LIT+ GQ  I+++DL++ YQT+ S GD YAVFVNAD+K R V+L + Y
Sbjct: 993  KATKAEVDRDVTLITQAGQDGIQQEDLIMGYQTVASNGDRYAVFVNADNKTRKVVLPQAY 1052

Query: 1114 KHLLKGQVIVDADQAGIKPISTPRGVHFEKDSLLIDPLTAIVIKV-GKVAPSPKEELQAD 1172
            ++LL  QV+VDA+QAG+  I+ P+GV F K+ L I+ LTA+V+KV  K A   +++ Q D
Sbjct: 1053 RYLLGAQVLVDAEQAGVTAIAKPKGVQFTKEGLTIEGLTALVLKVSSKTANPSQQKSQTD 1112

Query: 1173 YPKTQSFKESKTVEKVNRIANKTSITPVVSKKADSYLTNEANLPKTGDKSSKILSVVGIS 1232
              +T++   SK ++K      K + T                LPKTG+ SSK L   GI+
Sbjct: 1113 NHQTKTPDGSKDLDKSLMTRPKRAKT-------------NQKLPKTGEASSKGLLAAGIA 1159

Query: 1233 ILASLLALVGLSLKRNR 1249
            +   LL  + L +KR +
Sbjct: 1160 L---LLLAISLLMKRQK 1173
```

A related GBS gene <SEQ ID 8673> and protein <SEQ ID 8674> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:     -0.88
GvH: Signal Score (-7.5): 4.13
     Possible site: 41
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 3 value: -10.08 threshold:  0.0
   INTEGRAL    Likelihood =-10.08   Transmembrane 1225 -1241 (1222 -1247)
   INTEGRAL    Likelihood = -2.44   Transmembrane   19 -  35 (  18 -  36)
   INTEGRAL    Likelihood = -0.11   Transmembrane 1146 -1162 (1146 -1162)
   PERIPHERAL  Likelihood =  2.44     653
 modified ALOM score:   2.52

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5034(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 1081-1085
```

The protein has homology with the following sequences in the databases:

```
ORF00953(1111 - 3768 of 4356)
EGAD|165156|TM1845(18 - 840 of 843) pullulanase {Thermotoga maritima}SP|O33840|PULA_THEMA
PULLULANASE PRECURSOR (EC 3.2.1.41) (ALPHA
-DEXTRIN           ENDO-1,6-ALPHA-           GLUCOSIDASE)           (PULLULAN          6-
GLUCANOHYDROLASE).GP|2815006|emb|CAA04522.1||AJ001087 pullulanase {Thermotoga mari
tima}GP|4982428|gb|AAD36907.1|AE001821_7|AE001821          pullulanase          {Thermotoga
maritima}PIR|H72204|H72204 pullulanase - Thermotoga mariti
ma (strain MSB8)
%Match = 8.4
%Identity = 30.6  %Similarity = 52.8
Matches = 210  Mismatches = 298  Conservative Sub.s = 152

1032      1062      1092      1122      1152      1182      1212      1242
NKAGTNLSGDHHIPLLRPEMNQVWIDEKYGTHTYQPLKEGYVRINYLSSSSNYDHLSAWLFKDVATPSTTWPDGSNFVNQ
                   |            :   : |  : ::|      ||  : |::     | :       |  :
                   MKTKLWLLLVLLLSALIFSETTIVVHYHRYDGKYDGWNLWIWP--VEPVSQEGKAYQFTGE
                            10        20        30        40          50

1272      1302      1329      1359                1668      1698
GLYGRYIDVSLKTNAKEIGFLI-LDESKTGDAVKVQPNDYVFRDLA~~~~PKQGHFNISYNGNNVMTRQSWEFKDQL---
  :|:   | |  :  ::| :: |:|                                            |: ||
DDFGKVAVVKLPMDLTKVGIIVRLNE--------------------~~~~--------------------WQAKDVAKDR
```

```
         70        80                                                    90

1746      1776      1806      1836
-----------------~~~~--------------------YAYSGNLGAVLNQDGSKVEASLWSPSADSVTMIIYDKDN
                                         | | | |||| : :  |    :||| :  | ::::
FIEIKDGKAEVWILQGV~~~~ELIIEGYKPARVIMMEILDDYYYDGELGAVYSPE--KTIFRVWSPVSKWVKVLLFKNGE
            110               210       220       230         240       250

1866      1896      1926      1956      1986      2016      2046      2076
QNRVVATTPLMKNNKGVWQTILDTKLGIKNYTGYYYLYEIKRGKDKVKILDPYAKSLAEWDSNTVNDDIKTAKAAFVNPS
         :     |||: :::  |      | :|||:::        :|||:|::             : |:| || :
DTEPYQVVNMEYKGNGVWEAVVEGDL-----DGVFYLYQLENYGKIRTTVDPYSKAVYA----------NSKKSAVVNLA
            270       280            290       300                 310       320

2106      2136      2166      2196      2226                2253      2283
QLGPQNLSFAKIANFKGRQDAVIYEAHVRDFTSDRSLDGKLKNQFGTFAAFSEK-----------LDYLQKLGVTHIQLL
:  |:     :    :| :||:||| |: | |     :  :||: | :  ::|:           | :| :|||||:::|
RTNPEGWENDRGPKIEGYEDAIIYEIHIADITG--LENSGVKNK-GLYLGLTEENTKGPGGVTTGLSHLVELGVTHVHIL
       330       340       350         360        370       380       390

2313      2343      2373      2403      2433      2463      2493
PVLSYFYVNEMDKSRSTAYTSSDNNYNWGYDPQSYFALSGMYSEKPKDPSARIAELKQLIHDIHKRGMGVILDVVYNHT-
| : ::  :|:||         :  |||||||  :    | ||  ||:|  || |:|:::  :|| |:|||:|:|: ||
PFFDFYTGDELDK-------DFEKYYNWGYDPYLFMVPEGRYSTDPKNPHTRIREVKEMVKALHKHGIGVIMDMVFPHTY
          410              420       430       440       450       460       470

2544      2574      2601      2631      2661      2691      2721      2751
--AKTYLFEDIEPNYYHFMNEDGSP-RESFGGGRLGTTHAMSRRVLVDSIKYLTSEFKVDGFRFDMMGDHDAAAIELAYK
   :   |:   | |:: ::: |:   ||  |  : :   | |: :||:: |   |: :|||||| ||  |   :    :
GIGELSAFDQTVPYYFYRIDKTGAYLNESGCGNVIASERPMMRKFIVDTVTYWVKEYHIDGFRFDQMGLIDKKTMLEVER
       480       490       500       510       520       530       540       550

2781      2811      2841      2871      2901      2931                2979
EAKAINPNMIMIGEGWRTFQGDQGQPVKPADQDWMKSTDTVGVFSDDIRNSLKSGFPNEGTPAFITGG----PQSLQGIF
    |:| :|: || |    |  | |::    |   :   |  |:|: |::::    |     |: ||     :  :|:
ALHKIDPTIILYGEPW----GGWGAPIRFGKSD--VAGTHVAAFNDEFRDAIRGSVFNPSVKGFVMGGYGKETKIKRGVV
       560           570        580       590        600       610       620

3030      3060      3084      3114      3144      3174      3204
KNIKAQPG---NFEADSPGDVVQYIAAHDNLTLHD--VIAKSINKDPKVAEEEIHRRLRLGNVMILTSQGTAFIHSGQEY
 :|        :|  | | : : | | ||| || |   :|   :|  :  |||:    :|   ::|||||  |:| ||::
GSINYDGKLIKSFALD-PEETINYAACHDNHTLWDKNYLAAKADKKKEWTEEELKNAQKLAGAILLTSQGVPFLHGGQDF
             640        650       660       670       680       690       700

3234      3264      3294      3324      3354      3384      3414      3444
GRTKRLLNPDYMTKVSDDKLPNKATLIEAVKEYPYFIHDSYDSSDAINHFDWAAATDNNKHPISTKTQAYTAGLITLRRS
 |||                               |  :||::  :|| ||:       :         |  ||| ||:
CRTKN------------------------------FNDNSYNAPISINGFDY------ERKLQFIDVFNYHKGLIKLRKE
                                  710       720             730       740

3474                3504      3534      3564      3594      3624      3654
TDAFRKLSKAEI----------DREVSLITEVGQGDIKEKDLVIAYQTIDSKGDIYAVFVNADSKARNVLLGEKYKHLLK
  |||  :  ||           | |::: :   |    ||:|: |                       |    || | :
HPAFRLKNAEEIKKHLEFLPGGRRIVAFMLKDHAGGDPWKDIVVIYN--------------------GNLEKTTYK-LPE
          760       770       780       790                           800

3678      3708      3738      3768      3798      3828      3858      3888
GQ--VIVDADQAGIKPISTPRGVHFEKDSLLIDPLTAIVIKVGKVAPSPKEELQADYPKTQSFKESKTVEKVNRIANKTS
|:  |:|:: :|| : | |  |      ::  :|||:| |:
GKWNVVVNSQKAGTEVIETVEG------TIELDPLSAYVLYRE
           820             830       840
```

SEQ ID 2598 (GBS5) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 7; MW 134kDa).

The His-fusion protein was purified as shown in Figure 190, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 860

A DNA sequence (GBSx0912) was identified in *S.agalactiae* <SEQ ID 2601> which encodes the amino acid sequence <SEQ ID 2602>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.72   Transmembrane  231 - 247 ( 228 - 251)
   INTEGRAL    Likelihood = -8.39   Transmembrane   50 -  66 (  44 -  68)
   INTEGRAL    Likelihood = -6.74   Transmembrane   23 -  39 (  20 -  41)
   INTEGRAL    Likelihood = -5.84   Transmembrane  173 - 189 ( 168 - 196)
   INTEGRAL    Likelihood = -4.41   Transmembrane  299 - 315 ( 297 - 318)
   INTEGRAL    Likelihood = -4.14   Transmembrane  115 - 131 ( 114 - 133)
   INTEGRAL    Likelihood = -3.35   Transmembrane   80 -  96 (  79 -  97)
   INTEGRAL    Likelihood = -0.48   Transmembrane   97 - 113 (  97 - 113)

----- Final Results -----
              bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8675> which encodes amino acid sequence <SEQ ID 8676> was also identified. Analysis of this protein sequence reveals the following:

```
SRCFLG: 0
McG: Length  of  UR:   19
     Peak Value of UR:   3.08
     Net Charge of CR: 1
McG: Discrim Score:      9.76
GvH: Signal Score (-7.5): -4.57
     Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 7 value: -10.72 threshold:  0.0
   INTEGRAL    Likelihood =-10.72   Transmembrane  217 - 233 ( 214 - 237)
   INTEGRAL    Likelihood = -8.39   Transmembrane   36 -  52 (  30 -  54)
   INTEGRAL    Likelihood = -6.74   Transmembrane    9 -  25 (   6 -  27)
   INTEGRAL    Likelihood = -5.84   Transmembrane  159 - 175 ( 154 - 182)
   INTEGRAL    Likelihood = -4.14   Transmembrane  101 - 117 ( 100 - 119)
   INTEGRAL    Likelihood = -3.35   Transmembrane   66 -  82 (  65 -  83)
   INTEGRAL    Likelihood = -0.48   Transmembrane   83 -  99 (  83 -  99)
   PERIPHERAL  Likelihood =  0.26     136
 modified ALOM score:   2.64
icml HYPID: 7  CFP: 0.529

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB08178 GB:AB036768 exfoliative toxin A [Staphylococcus hyicus]
 Identities = 134/298 (44%), Positives = 197/298 (65%)

Query: 22  PLVMAGLVLGLLALGNLLEGYGTYVRYCLGLVALVFWIFLIKGILKNKKESRKELSNPLI 81
            PLV +GLVLGLL LGNLL+     +    G++A++ W+ L+  +  N    + +L++PL+
Sbjct: 7   PLVSSGLVLGLLGLGNLLKDVSLSLNALCGILAILVWLHLLYSMFNNVNHVKNQLNSPLV 66

Query: 82  ASVFTTFFMAGMILSTYILLFRSLGIWVAVLSKGVWWLSFIALIIHMAIFSWKYLRHFSM 141
```

```
           +SVFTTFFM+G + +TY+  F S   ++  L   +W L  I ++ HM IFS KYL+ FS+
Sbjct: 67  SSVFTTFFMSGFLGTTYLNTFFSHISFIHHLITPLWLLCLIGILTHMIIFSHKYLKSFSL 126

Query: 142 ANLFPSWSVLYVGIGVASLTAPISGQFTIGKIVFWYGFIATLVLLPFLFIKAYKIGLPSA 201
            N++PSW+VLY+GI +A LTAP+SG F IGK+   YGF+AT ++LP +F +      L ++
Sbjct: 127 ENVYPSWTVLYIGIAIAGLTAPVSGYFFIGKLTVIYGFVATCIVLPLVFKRLKTYPLQTS 186

Query: 202 VKPNITTICAPMSLITAGYVNSFVSPNRGLLLLLIVMAQFLYFFILFQVPKLLIGDFTPG 261
           +KPN +TICAP SL+ A YV +F   +  +++L ++++Q  YF+I+FQ+PKLL   F+P
Sbjct: 187 IKPNTSTICAPFSLVAAAYVLAFPEAHDFVVILFLILSQVFYFYIVFQLPKLLREPFSPV 246

Query: 262 FSAFTFPLVISATSLKLSIQHLSLPVDIQGLVHFEIGTTTLIVMIVMVRYIFFLRRTI 319
           FSAFTFPLVISAT+LK S+  L  P    GL+ FE    T+IV  V   YI    + +
Sbjct: 247 FSAFTFPLVISATALKNSMPILIFPEIWNGLLMFETVLATVIVFRVFFGYIHLFLKPV 304
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2603> which encodes the amino acid sequence <SEQ ID 2604>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.82   Transmembrane  169 - 185 ( 163 - 189)
   INTEGRAL    Likelihood = -8.49   Transmembrane   50 -  66 (  38 -  69)
   INTEGRAL    Likelihood = -7.86   Transmembrane  228 - 244 ( 224 - 247)
   INTEGRAL    Likelihood = -5.15   Transmembrane  288 - 304 ( 284 - 306)
   INTEGRAL    Likelihood = -3.29   Transmembrane  108 - 124 ( 107 - 126)
   INTEGRAL    Likelihood = -3.29   Transmembrane  140 - 156 ( 140 - 161)
   INTEGRAL    Likelihood = -1.33   Transmembrane   84 - 100 (  84 - 100)

----- Final Results -----
               bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 138/305 (45%), Positives = 200/305 (65%), Gaps = 5/305 (1%)

Query: 12  RYMMKNWEKPPLVMAGLVLGLLALGNLLEGYGTYVRYCLGLVALVFWIFLIKGILKNKKE 71
           R +MK+ + PPLVM+GL LG L+ GNLL  Y +   Y   L AL  +  L+ G+++N  +
Sbjct: 12  RTLMKHLKTPPLVMSGLALGTLSFGNLLATYVSIFNYLGILAALFIYGILLVGMVRNLND 71

Query: 72  SRKELSNPLIASVFTTFFMAGMILSTYILLFRSLGIWVAVLSKGVWWLSFIALIIHMAIF 131
           ++ +L  PLIASVF TFFM GM+LS+  L     G W+  L+   WWL F+  ++ +A +
Sbjct: 72  TKMQLRQPLIASVFPTFFMTGMLLSSLFLKVTG-GCWLGFLT---WWLFFLGNLVLIAYY 127

Query: 132 SWKYLRHFSMANLFPSWSVLYVGIGVASLTAPISGQFTIGKIVFWYGFIATLVLLPFLFI 191
            ++++  FS  N+FPSWSVL+VGI +A+LTAP S QF +G+++FW   + T V+LPF+
Sbjct: 128 QYRFVFSFSWDNVFPSWSVLFVGIAMAALTAPASRQFLLGQVIFWVCLLLTAVILPFMAK 187

Query: 192 KAYKIGLPSAVKPNITTICAPMSLITAGYVNSFVSPNRGLLLLLIVMAQFLYFFILFQVP 251
           K Y IGL  AV PNI+T CAP+SL++A Y+ +F  P  G+++ L+V +Q LY F++ Q+P
Sbjct: 188 KTYGIGLGQAVMPNISTFCAPLSLLSASYLATFPRPQVGMVIFLLVSSQLLYAFVVVQLP 247

Query: 252 KLLIGDFTPGFSAFTFPLVISATSLKLSIQHLSLP-VDIQGLVHFEIGTTTLIVMIVMVR 310
           +LL   F PGFSAFTFP VISATSLK+++  L    +  Q L+  E+    T +V  V
Sbjct: 248 RLLNRPFNPGFSAFTFPFVISATSLKMTLSFLGWQGLGWQVLLLGEVLLATALVTYVYGA 307

Query: 311 YIFFL 315
           Y+ FL
Sbjct: 308 YLRFL 312
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 861

A DNA sequence (GBSx0913) was identified in *S.agalactiae* <SEQ ID 2605> which encodes the amino acid sequence <SEQ ID 2606>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2607> which encodes the amino acid sequence <SEQ ID 2608>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 45/57 (78%), Positives = 53/57 (92%)

Query: 1  MVKKFAFAKGIATGVVATAATLAGAAFAIKKTIIEPEEEKIAFIEENRKKAARKRVS 57
           MVKK+ F KG+ATGV+ATAAT+AGA FA+KKTII+PEEEK AFIEENRKKAAR+RV+
Sbjct: 1  MVKKYQFVKGLATGVLATAATVAGAVFAVKKTIIDPEEEKAAFIEENRKKAARRRVA 57
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 862

A DNA sequence (GBSx0914) was identified in *S.agalactiae* <SEQ ID 2609> which encodes the amino acid sequence <SEQ ID 2610>. This protein is predicted to be tRNA isopentenylpyrophosphate transferase (miaA). Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9897> which encodes amino acid sequence <SEQ ID 9898> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06085 GB:AP001515 tRNA isopentenylpyrophosphate transferase
           [Bacillus halodurans]
 Identities = 139/311 (44%), Positives = 200/311 (63%), Gaps = 21/311 (6%)

Query: 7   KIKLIAVVGPTAVGKTALGIELAKTFNGEIISGDSQQVYQKLDIGTAKASKEEQEQAYHH 66
            K KL+A+VGPTAVGKT    + LAK  NGE+ISGDS QVY+ +DIGTAK + EE +    HH
Sbjct: 2   KEKLVAIVGPTAVGKTKTSVMLAKRLNGEVISGDSMQVYRGMDIGTAKITAEEMDGVPHH 61
```

```
Query: 67  LIDVREVNENYSVYDFVKEAKVAIDTIISKGKIPIIVGGTGLYLQSLFEGYHLGGEVNQE 126
           LID+++ +E++SV DF   A   I  I  +G++P +VGGTGLY+ ++   ++LG     E
Sbjct: 62  LIDIKDPSESFSVADFQDLATPLITEIHERGRLPFLVGGTGLYVNAVIHQFNLGDIRADE 121

Query: 127 TLMAYREKLE----SLSDEDLFEKLT----EQSIIIPQVNRRRAIRALELAKF------- 171
               YR +LE    S   + L +KL+    + +  I   N RR IRALE+ K
Sbjct: 122 D---YRHELEAFVNSYGVQALHDKLSKIDPKAAAAIHPNNYRRVIRALEIIKLTGKTVTE 178

Query: 172 -GNDLQNSESPYDVLLIGLNDDRQVLYDRINRRVDLMMDNGLLDEAKWLYD-NYPSVQAS 229
                + + SPY++++IGL  +R VLYDRINRRVD M++ GL+DEAK LYD      Q+
Sbjct: 179 QARHEEETPSPYNLVMIGLTMERDVLYDRINRRVDQMVEEGLIDEAKKLYDRGIRDCQSV 238

Query: 230 KGIGYKELFPYFSKQIPLEEAVDKLKQNTRRFAKRQLTWFRNRMNVEFIMVGEENYQQKI 289
           + IGYKE++ Y    + LEEA+D LK+N+RR+AKRQLTWFRN+ NV +  + + ++ +KI
Sbjct: 239 QAIGYKEMYDYLDGNVTLEEAIDTLKRNSRRYAKRQLTWFRNKANVTWFDMTDVDFDKKI 298

Query: 290 KRKVSDFLSSK 300
             ++ +F++ K
Sbjct: 299 -MEIHNFIAGK 308
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2611> which encodes the amino acid sequence <SEQ ID 2612>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 202/296 (68%), Positives = 250/296 (84%)

Query: 5   MRKIKLIAVVGPTAVGKTALGIELAKTFNGEIISGDSQQVYQKLDIGTAKASKEEQEQAY 64
           M KIK++ +VGPTAVGKTALGI LAK FNGEIISGDSQQVY++LDIGTAKA++EEQE A
Sbjct: 1   MTKIKIVVIVGPTAVGKTALGISLAKAFNGEIISGDSQQVYRQLDIGTAKATQEEQEAAV 60

Query: 65  HHLIDVREVNENYSVYDFVKEAKVAIDTIISKGKIPIIVGGTGLYLQSLFEGYHLGGEVN 124
           HHLID+REV E+YS YDFV++A+ +I  I+S+GK+PIIVGGTGLYLQSL EGYHLGG+V+
Sbjct: 61  HHLIDIREVTESYSAYDFVQDAQKSISDIVSRGKLPIIVGGTGLYLQSLLEGYHLGGQVD 120

Query: 125 QETLMAYREKLESLSDEDLFEKLTEQSIIIPQVNRRRAIRALELAKFGNDLQNSESPYDV 184
           QE + AYR +LE L D DL+E+L   +I I QVNRRRAIRALELA+F ++L+N+E+ Y+
Sbjct: 121 QEAVKAYRNELEQLDDHDLYERLQVNNITIEQVNRRRAIRALELAQFADELENAETAYEP 180

Query: 185 LLIGLNDDRQVLYDRINRRVDLMMDNGLLDEAKWLYDNYPSVQASKGIGYKELFPYFSKQ 244
           L+IGLNDDRQV+YDRIN+RV+ M++NGLL+EAKWLY++YP+VQAS+GIGYKELFPYF  +
Sbjct: 181 LIIGLNDDRQVIYDRINQRVNRMIENGLLEEAKWLYEHYPTVQASRGIGYKELFPYFVGE 240

Query: 245 IPLEEAVDKLKQNTRRFAKRQLTWFRNRMNVEFIMVGEENYQQKIKRKVSDFLSSK 300
           + L EA D+LKQNTRRFAKRQLTWFRNRM V F  +   +Y Q +  +V DFL  K
Sbjct: 241 MTLAEASDQLKQNTRRFAKRQLTWFRNRMAVSFTAITAPDYPQVVHDRVRDFLGQK 296
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 863

A DNA sequence (GBSx0915) was identified in *S.agalactiae* <SEQ ID 2613> which encodes the amino acid sequence <SEQ ID 2614>. This protein is predicted to be hflX (hflX). Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06081 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 182/406 (44%), Positives = 254/406 (61%), Gaps = 12/406 (2%)

Query: 9   ERVILVGVELQDT--ENFEMSMEELASLAKTAGANVVNHYYQKRDKYDSKSFIGSGKLEE 66
           ERV LV  +L +   E FE S+EEL +L  TA   V++   QKR+  +  ++IG GKL+E
Sbjct: 10  ERVFLVACQLPNMTDEQFEASLEELEALTLTAQGTVIDRLTQKREAIEPATYIGRGKLDE 69

Query: 67  IKAIVEADEIDTVVVNNRLTPRQNSNLEAELGVKVIDRMQLILDIFAMRARSHEGKLQVH 126
           +   +E  E D V+VN  L+  Q  NL   LGV+VIDR QLILDIFA RA+S EGKLQV
Sbjct: 70  LAIKMEEQEADLVIVNGELSGSQVRNLTNRLGVRVIDRTQLILDIFAGRAKSREGKLQVE 129

Query: 127 LAQLKYMLPRLVGQGIMLSRQAGGIGSRGPGESQLELNRRSIRHQISDIERQLKIVEKNR 186
           LAQL Y+LPR+VGQG  LSR  GGIG+RGPGE++LE +RR IR +++DI++QLK   K+R
Sbjct: 130 LAQLNYLLPRIVGQGQGLSRLGGGIGTRGPGETKLETDRRHIRKRMADIDKQLKHTVKHR 189

Query: 187 ETVRERRVDSTTFKIGLIGYTNAGKSTIMNVLTDDKQYEANELFATLDATTKQIYLQNQF 246
           +  R RR  + TF+I L+GYTNAGKST++N LT    YE + LFATLD  T+++ L +
Sbjct: 190 DRYRARRERNQTFRIALVGYTNAGKSTLLNRLTASDSYEEDLLFATLDPMTRKMRLPSGM 249

Query: 247 QVTLTDTVGFIQDLPTELVAAFKSTLEESRHVDLLFHVIDASDPNHEEHEKVVMEILKDL 306
           +V L+DTVGFI  LPT LVAAF+STLEE +H DLL HV+D S    + H + V E+L  L
Sbjct: 250 EVILSDTVGFINQLPTTLVAAFRSTLEEVKHADLLLHVVDRSSEQLQAHMETVSELLHQL 309

Query: 307 DMIDIPRLAIYNKMDVTEQLNATTFP-----NVRIAAKKQGSKDLLRRLIVDEIRHIFDE 361
           ++     L +YNK D   + N    P      + ++A K+     LR++I   +  +F
Sbjct: 310 EVDQSQMLVVYNKAD---KPNLPIIPVHQQNGIEMSAHKREDIQRLRQMIERTLVDLFTP 366

Query: 362 FSIRVHQNQAYKLYDLNKIALLDTYTFEEEYE--NITGYISPKQKW 405
           +   +  ++  KL  L +  ++    ++E+ E   + GY+ P   W
Sbjct: 367 YVTELASDEGNKLAKLRRETIMTEMKWDEDRECYQVKGYVHPNHAW 412
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2615> which encodes the amino acid sequence <SEQ ID 2616>. Analysis of this protein sequence reveals the following:

```
     Possible site: 48

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB06081 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 185/403 (45%), Positives = 246/403 (60%), Gaps = 6/403 (1%)

Query: 13  ERVILLGVEL--QTTEHFDMSMTELANLAKTAGVKVMASFSQKRERYDSKTFIGSGKLDE 70
           ERV L+  +L   T E F+ S+ EL  L  TA   V+   +QKRE  +  T+IG GKLDE
Sbjct: 10  ERVFLVACQLPNMTDEQFEASLEELEALTLTAQGTVIDRLTQKREAIEPATYIGRGKLDE 69

Query: 71  IKAIVEADEIDAVIVNNRLTARQNANLEAVLEVKVIDRMQLILDIFAMRARSHEGKLQVH 130
           +   +E  E D VIVN  L+  Q  NL   L V+VIDR QLILDIFA RA+S EGKLQV
Sbjct: 70  LAIKMEEQEADLVIVNGELSGSQVRNLTNRLGVRVIDRTQLILDIFAGRAKSREGKLQVE 129

Query: 131 LAQLKYMLPRLVGQGIMLSRQAGGIGSRGPGESQLELNRRSIRHQIADIERQLTQVEKNR 190
```

```
           LAQL Y+LPR+VGQG  LSR  GGIG+RGPGE++LE +RR IR ++ADI++QL    K+R
Sbjct: 130 LAQLNYLLPRIVGQGQGLSRLGGGIGTRGPGETKLETDRRHIRKRMADIDKQLKHTVKHR 189

Query: 191 QTIRDRRVGSDTFKIGLIGYTNAGKSTIMNLLTDDSHYEANELFATLDATTKQLYLENQF 250
              R RR  + TF+I L+GYTNAGKST++N LT    YE + LFATLD  T+++ L +
Sbjct: 190 DRYRARRERNQTFRIALVGYTNAGKSTLLNRLTASDSYEEDLLFATLDPMTRKMRLPSGM 249

Query: 251 QATLTDTVGFIQDLPTELVAAFKSTLEESKYVDLLLHVIDASDPNHSEQEKVVLNLLKEL 310
           +  L+DTVGFI  LPT LVAAF+STLEE K+ DLLLHV+D S         + V  LL +L
Sbjct: 250 EVILSDTVGFINQLPTTLVAAFRSTLEEVKHADLLLHVVDRSSEQLQAHMETVSELLHQL 309

Query: 311 DMLNIPRLAIYNKVDIAEQ--FTATAFPNIRISARSKDSKILLRRLIIDQIRDQFVPFRI 368
           ++     L +YNK D              I +SA ++     LR++I   + D F P+
Sbjct: 310 EVDQSQMLVVYNKADKPNLPIIPVHQQNGIEMSAHKREDIQRLRQMIERTLVDLFTPYVT 369

Query: 369 KVHQDKAYKLYDLNRVALLDHYTFDQEIE--DISGYISPKQQW 409
           ++  D+  KL  L R  ++    +D++ E   + GY+ P   W
Sbjct: 370 ELASDEGNKLAKLRRETIMTEMKWDEDRECYQVKGYVHPNHAW 412
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 326/412 (79%), Positives = 375/412 (90%)

Query: 1   MIETKEEQERVILVGVELQDTENFEMSMEELASLAKTAGANVVNHYYQKRDKYDSKSFIG 60
           MIETK +QERVIL+GVELQ TE+F+MSM ELA+LAKTAG  V+  + QKR++YDSK+FIG
Sbjct: 5   MIETKRQQERVILLGVELQTTEHFDMSMTELANLAKTAGVKVMASFSQKRERYDSKTFIG 64

Query: 61  SGKLEEIKAIVEADEIDTVVVNNRLTPRQNSNLEAELGVKVIDRMQLILDIFAMRARSHE 120
           SGKL+EIKAIVEADEID V+VNNRLT RQN+NLEA L VKVIDRMQLILDIFAMRARSHE
Sbjct: 65  SGKLDEIKAIVEADEIDAVIVNNRLTARQNANLEAVLEVKVIDRMQLILDIFAMRARSHE 124

Query: 121 GKLQVHLAQLKYMLPRLVGQGIMLSRQAGGIGSRGPGESQLELNRRSIRHQISDIERQLK 180
           GKLQVHLAQLKYMLPRLVGQGIMLSRQAGGIGSRGPGESQLELNRRSIRHQI+DIERQL
Sbjct: 125 GKLQVHLAQLKYMLPRLVGQGIMLSRQAGGIGSRGPGESQLELNRRSIRHQIADIERQLT 184

Query: 181 IVEKNRETVRERRVDSTTFKIGLIGYTNAGKSTIMNVLTDDKQYEANELFATLDATTKQI 240
            VEKNR+T+R+RRV S TFKIGLIGYTNAGKSTIMN+LTDD  YEANELFATLDATTKQ+
Sbjct: 185 QVEKNRQTIRDRRVGSDTFKIGLIGYTNAGKSTIMNLLTDDSHYEANELFATLDATTKQL 244

Query: 241 YLQNQFQVTLTDTVGFIQDLPTELVAAFKSTLEESRHVDLLFHVIDASDPNHEEHEKVVM 300
           YL+NQFQ TLTDTVGFIQDLPTELVAAFKSTLEES++VDLL HVIDASDPNH E EKVV+
Sbjct: 245 YLENQFQATLTDTVGFIQDLPTELVAAFKSTLEESKYVDLLLHVIDASDPNHSEQEKVVL 304

Query: 301 EILKDLDMIDIPRLAIYNKMDVTEQLNATTFPNVRIAAKKQGSKDLLRRLIVDEIRHIFD 360
            +LK+LDM++IPRLAIYNK+D+ EQ  AT FPN+RI+A+ + SK LLRRLI+D+IR  F
Sbjct: 305 NLLKELDMLNIPRLAIYNKVDIAEQFTATAFPNIRISARSKDSKILLRRLIIDQIRDQFV 364

Query: 361 EFSIRVHQNQAYKLYDLNKIALLDTYTFEEEYENITGYISPKQKWKLEEFYD 412
            F I+VHQ++AYKLYDLN++ALLD YTF++E E+I+GYISPKQ+W+L++FY+
Sbjct: 365 PFRIKVHQDKAYKLYDLNRVALLDHYTFDQEIEDISGYISPKQQWRLDDFYE 416
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 864

A DNA sequence (GBSx0916) was identified in *S.agalactiae* <SEQ ID 2617> which encodes the amino acid sequence <SEQ ID 2618>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2044(Affirmative) < succ>
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
            bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2619> which encodes the amino acid sequence <SEQ ID 2620>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3436(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 124/209 (59%), Positives = 150/209 (71%)

Query: 1   MIDYIDLALTYGGFTSLDKVYLEKKLDGLSKQQRLDFITPPPSVINAYFAEIYQKQGPEA 60
           M +YIDLA TYGGFTSLD  YL   L  L+ QQ+L FITPPPSVINAYFAEIYQKQ P+A
Sbjct: 5   MNNYIDLAKTYGGFTSLDTNYLNHLLASLTDQQKLAFITPPPSVINAYFAEIYQKQSPQA 64

Query: 61  ATDYYFDLSKALGLFPKHLSFDEEKPFIRLNLSGKSFGFAYLNDQEEASVFSEVKEVITP 120
           ATDYYF+LSKALGLF    SF+EEKPF+RLNLSGK++GFAY NDQE A VFSE  E   P
Sbjct: 65  ATDYYFNLSKALGLFTDQPSFEEEKPFVRLNLSGKAYGFAYQNDQEVALVFSEKAEPKKP 124

Query: 121 QLLLEIAQIFPQYKVYRDRSGIRMAKIDFDETESQNITPETSLLGNVLQLKKDIIKITSF 180
           +L  E+ QIFPQY VY D+  ++M    F++ E ++ITP+ +LL  + +L   I  + F
Sbjct: 125 ELFFELTQIFPQYMVYEDKGQLKMQAKQFEQGECEDITPDDTLLSKIYRLANGITMLKGF 184

Query: 181 NQEELLELVKTKSGKYYYSSQGRESVIYI 209
           N EEL  L +T SG+ YY    RE +IYI
Sbjct: 185 NVEELWALSQTFSGQKYYDFAQREFMIYI 213
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 865

A DNA sequence (GBSx0917) was identified in *S.agalactiae* <SEQ ID 2621> which encodes the amino acid sequence <SEQ ID 2622>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1060(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9895> which encodes amino acid sequence <SEQ ID 9896> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14316 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 156/309 (50%), Positives = 210/309 (67%), Gaps = 5/309 (1%)

Query: 1   MEIQFLGTGAGQPAKARNVSSLVLKLLDEINEVWMFDCGEGTQRQILETTIKPRKVKKIF 60
           ME+ FLGTGAG PAKARNV+S+ LKLL+E   VW+FDCGE TQ QIL TTIKPRK++KIF
Sbjct: 1   MELLFLGTGAGIPAKARNVTSVALKLLEERRSVWLFDCGEATQHQILHTTIKPRKIEKIF 60

Query: 61  ITHMHGDHVFGLPGFLSSRAFQANEEQTDLDIYGPVGIKSFVMTALRTSGSRLPYRIHFH 120
           ITHMHGDHV+GLPG L SR+FQ  E++  L +YGP GIK+F+ T+L  + + L Y +
```

```
Sbjct: 61  ITHMHGDHVYGLPGLLGSRSFQGGEDE--LTVYGPKGIKAFIETSLAVTKTHLTYPLAIQ 118

Query: 121 EFDESSLGKIMETDKFTVYAEKLDHTIFCMGYRVVQKDLEGTLDAEALKLAGVPFGPLFG 180
           E +E   G + E D+F V A  + H +   GYRV +KD+ G+L A+ LK   +P GP++
Sbjct: 119 EIEE---GIVFEDDQFIVTAVSVIHGVEAFGYRVQEKDVPGSLKADVLKEMNIPPGPVYQ 175

Query: 181 KVKNGENVTLEDGREIIAKDYISEPKKGKVITILGDTRKTDASIRLALGADVLVHESTYG 240
           K+K GE VTLEDGR I   D++  PKKG+ +   GDTR +D    LA   DVLVHE+T+
Sbjct: 176 KIKKGETVTLEDGRIINGNDFLEPPKKGRSVVFSGDTRVSDKLKELARDCDVLVHEATFA 235

Query: 241 KGDERIAKSHGHSTNMQAADIAKQANAKRLLLNHVSARFMGRDCWQMEEDAKTIFSNTHL 300
           K D ++A  + HST  QAA  AK+A AK+L+L H+SAR+ G    +++++A  +F N+
Sbjct: 236 KEDRKLAYDYYHSTTEQAAVTAKEARAKQLILTHISARYQGDASLELQKEAVDVFPNSVA 295

Query: 301 VRDLEEVGI 309
             D  EV +
Sbjct: 296 AYDFLEVNV 304
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2623> which encodes the amino acid sequence <SEQ ID 2624>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2352(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 253/307 (82%), Positives = 285/307 (92%)

Query: 1   MEIQFLGTGAGQPAKARNVSSLVLKLLDEINEVWMFDCGEGTQRQILETTIKPRKVKKIF 60
           ME+QFLGTGAGQPAK RNVSSL LKLLDEINEVWMFDCGEGTQRQILETTIKPRK++KIF
Sbjct: 1   MELQFLGTGAGQPAKQRNVSSLALKLLDEINEVWMFDCGEGTQRQILETTIKPRKIRKIF 60

Query: 61  ITHMHGDHVFGLPGFLSSRAFQANEEQTDLDIYGPVGIKSFVMTALRTSGSRLPYRIHFH 120
           ITH+HGDH+FGLPGFLSSR+FQA+EEQTDLDIYGP+GIK++V+T+L+ SG+R+PY+IHFH
Sbjct: 61  ITHLHGDHIFGLPGFLSSRSFQASEEQTDLDIYGPIGIKTYVLTSLKVSGARVPYQIHFH 120

Query: 121 EFDESSLGKIMETDKFTVYAEKLDHTIFCMGYRVVQKDLEGTLDAEALKLAGVPFGPLFG 180
           EFD+ SLGKIMETDKF VYAE+L HTIFCMGYRVVQKDLEGTLDAEALK AGVPFGPLFG
Sbjct: 121 EFDDKSLGKIMETDKFEVYAERLAHTIFCMGYRVVQKDLEGTLDAEALKAAGVPFGPLFG 180

Query: 181 KVKNGENVTLEDGREIIAKDYISEPKKGKVITILGDTRKTDASIRLALGADVLVHESTYG 240
           K+KNG++V LEDGR I AKDYIS PKKGK+ITI+GDTRKT AS++LA  ADVLVHESTYG
Sbjct: 181 KIKNGQDVELEDGRLICAKDYISAPKKGKIITIIGDTRKTSASVKLAKDADVLVHESTYG 240

Query: 241 KGDERIAKSHGHSTNMQAADIAKQANAKRLLLNHVSARFMGRDCWQMEEDAKTIFSNTHL 300
           KGDERIA++HGHSTNMQAA IA +A AKRLLLNHVSARF+GRDC QME+DA TIF N  +
Sbjct: 241 KGDERIARNHGHSTNMQAAQIAHEAGAKRLLLNHVSARFLGRDCRQMEKDAATIFENVKM 300

Query: 301 VRDLEEV 307
           V+DLEEV
Sbjct: 301 VQDLEEV 307
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 866

A DNA sequence (GBSx0918) was identified in *S.agalactiae* <SEQ ID 2625> which encodes the amino acid sequence <SEQ ID 2626>. This protein is predicted to be similar to ketoacyl reductase. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14310 GB:Z99116 similar to ketoacyl reductase [Bacillus subtilis]
 Identities = 100/253 (39%), Positives = 152/253 (59%), Gaps = 2/253 (0%)

Query: 3   RTILITGASGGLAQAIINQLPQDD-HLIVTGRSREKLEKLYGKRPNTLCLSLDITN-DNA 60
           + I ITGASGGL + I      +  H++++ R  ++L ++  K         +I   D
Sbjct: 7   KRIWITGASGGLGERIAYLCAAEGAHVLLSARREDRLIEIKRKITEEWSGQCEIFPLDVG 66

Query: 61  VTNMIEKIYGEFGQIDILINNAGFGSFKEFWDYSDEEVKDMFAVNTFATMSIARQIGHKM 120
               I ++  + G ID+LINNAGFG F+   D + +++K MF VN F  ++  + +  +M
Sbjct: 67  RLEDIARVRDQIGSIDVLINNAGFGIFETVLDSTLDDMKAMFDVNVFGLIACTKAVLPQM 126

Query: 121 SLVKSGHIVNIASMAGLIATSKASVYGASKFAVVGFSNALRLELAEKNVYVTSVNPGPIK 180
              K GHI+NIAS AG IAT K+S+Y A+K AV+G+SNALR+EL+   +YVT+VNPGPI+
Sbjct: 127 LEQKKGHIINIASQAGKIATPKSSLYSATKHAVLGYSNALRMELSGTGIYVTTVNPGPIQ 186

Query: 181 TGFFAQADPSGDYLASIGRFALTPEKVSKKVVSILGKNKRELNLPFILAFAHKYYSLFPK 240
           T FF+ AD  GDY  ++GR+ L P+ V+ ++ + +   KRE+NLP ++    K Y LFP
Sbjct: 187 TDFFSIADKGGDYAKNVGRWMLDPDDVAAQITAAIFTKKREINLPRLMNAGTKLYQLFPA 246

Query: 241 TADYFARKVFNYK 253
             +  A +    K
Sbjct: 247 LVEKLAGRALMKK 259
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2627> which encodes the amino acid sequence <SEQ ID 2628>. Analysis of this protein sequence reveals the following:

```
     Possible site: 18

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB05225 GB:AP001512 oxidoreductase [Bacillus halodurans]
 Identities = 107/259 (41%), Positives = 156/259 (59%), Gaps = 5/259 (1%)

Query: 1   MAQRIIVITGASGGLAQAIVKQLPKEDSLI-LLGRNKERLEHCYQHI----DNKECLELD 55
           M ++ I ITGAS GL + +      E++++ L  R++ERLE+  + +           +D
Sbjct: 1   MRKKTIFITGASSGLGRQLAIDFSWEETVLCLFARSQERLENVQRIVVENGGEAHIYPVD 60

Query: 56  ITNPVAIEKMVAQIYQRYGRIDVLINNAGYGAFKGFEEFSAQEIADMFQVNTLASIHFAC 115
           + +P +I++  A+     G +DVLINNAGYG F+ F +    E   MF+VN    +
Sbjct: 61  LADPQSIDRSFAEAISAVGVVDVLINNAGYGVFEPFCDSQMDENERMFRVNVFGLMRATA 120

Query: 116 LIGQKMAEQGQGHLINIVSMAGLIASAKSSIYSATKFALIGFSNALRLELADKGVYVTTV 175
            +   M EQG GH+INI S AG IA+AKS+IYSATK A++GF+N+LR+EL   G++V+ V
```

```
Sbjct: 121 AVLPTMREQGSGHIINIASQAGKIATAKSAIYSATKHAVLGFTNSLRMELKGTGIHVSAV 180

Query: 176 NPGPIATKFFDQADPSGHYLESVGKFTLQPNQVAKRLVSIIGKNKRELNLPFSLAVTHQF 235
           NPGPI T FFDQAD  G Y   V +  L P  V++++V +  K KRELNLP+ + +
Sbjct: 181 NPGPIQTPFFDQADKEGAYTSKVQRIMLDPEDVSEKIVQLTKKPKRELNLPWWMNIGATA 240

Query: 236 YTLFPKLSDYLARKVFNYK 254
           Y + P+L + LA K F  K
Sbjct: 241 YQVAPRLLELLAGKQFRQK 259
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 155/251 (61%), Positives = 200/251 (78%)

Query: 3   RTILITGASGGLAQAIINQLPQDDHLIVTGRSREKLEKLYGKRPNTLCLSLDITNDNAVT 62
           R I+ITGASGGLAQAI+ QLP++D LI+ GR++E+LE  Y    N  CL LDITN  A+
Sbjct: 4   RIIVITGASGGLAQAIVKQLPKEDSLILLGRNKERLEHCYQHIDNKECLELDITNPVAIE 63

Query: 63  NMIEKIYGEFGQIDILINNAGFGSFKEFWDYSDEEVKDMFAVNTFATMSIARQIGHKMSL 122
            M+ +IY  +G+ID+LINNAG+G+FK F ++S +E+ DMF VNT A++  A  IG KM+
Sbjct: 64  KMVAQIYQRYGRIDVLINNAGYGAFKGFEEFSAQEIADMFQVNTLASIHFACLIGQKMAE 123

Query: 123 VKSGHIVNIASMAGLIATSKASVYGASKFAVVGFSNALRLELAEKNVYVTSVNPGPIKTG 182
              GH++NI SMAGLIA++K+S+Y A+KFA++GFSNALRLELA+K VYVT+VNPGPI T
Sbjct: 124 QGQGHLINIVSMAGLIASAKSSIYSATKFALIGFSNALRLELADKGVYVTTVNPGPIATK 183

Query: 183 FFAQADPSGDYLASIGRFALTPEKVSKKVVSILGKNKRELNLPFILAFAHKYYSLFPKTA 242
           FF QADPSG YL S+G+F L P +V+K++VSI+GKNKRELNLPF LA  H++Y+LFPK +
Sbjct: 184 FFDQADPSGHYLESVGKFTLQPNQVAKRLVSIIGKNKRELNLPFSLAVTHQFYTLFPKLS 243

Query: 243 DYFARKVFNYK 253
           DY ARKVFNYK
Sbjct: 244 DYLARKVFNYK 254
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 867

A DNA sequence (GBSx0919) was identified in *S.agalactiae* <SEQ ID 2629> which encodes the amino acid sequence <SEQ ID 2630>. This protein is predicted to be single-stranded-DNA-specific exonuclease (recJ). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16   Transmembrane  197 - 213 ( 197 - 213)

----- Final Results -----
               bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB14721 GB:Z99118 similar to single-strand DNA-specific
          exonuclease [Bacillus subtilis]
 Identities = 276/772 (35%), Positives = 447/772 (57%), Gaps = 45/772 (5%)

Query: 1   MISAKYSWVLNNQKPDAGFFEASKKE-KISEAVASLIYSRGIKTSAELHHFLQTNLENLH 59
           M+++K  W +  Q+PD    ++  ++  I+  VASL+  RG  T+     FL T   + +
Sbjct: 1   MLASKMRWEI--QRPDQDKVKSLTEQLHITPLVASLLVKRGFDTAESARLFLHTKDADFY 58

Query: 60  DPYLLNDMDKAVNRIRRAIENNETILVYGDYDADGMTSASIMKEALDMMGAEVQVYLPNR 119
           DP+ +  M +A +RI++AI   E I++YGDYDADG+TS S+M   L  + A+V  Y+P+R
Sbjct: 59  DPFEMKGMKEAADRIKQAISQQEKIMIYGDYDADGVTSTSVMLHTLQKLSAQVDFYIPDR 118
```

```
Query: 120 FTDGYGPNQSVYKYFIEQQDVSLIITVDNGVAGHEAITYAQNQGVDVVVTDHHSMPADLP 179
           F +GYGPN+  ++  I+++  SLIITVD G+A       A+  G+DV++TDHH    +LP
Sbjct: 119 FKEGYGPNEQAFRS-IKERGFSLIITVDTGIAAVHEAKVAKELGLDVIITDHHEPGPELP 177

Query: 180 CAYAIIHPEHPDANYPFPYLAGCGVAFKVACALLETIPTEMLDLVAIGTIADMVSLTDEN 239
              AI+HP+ P   YPF  LAG GVAFK+A ALL  +P E+LDL AIGTIAD+V L DEN
Sbjct: 178 DVRAIVHPKQPGCTYPFKELAGVGVAFKLAHALLGELPDELLDLAAIGTIADLVPLHDEN 237

Query: 240 RIMVKAGLEVMKDSERIGLQELISLSNIDLKTLNEETIGFKIAPQLNALGRLDDPNPAIE 299
           R++   GLE ++ + R+GL+ELI LS  D+   NEET+GF++AP+LNA+GR++  +PA+
Sbjct: 238 RLIATLGLERLRRTNRLGLKELIKLSGGDIGEANEETVGFQLAPRLNAVGRIEQADPAVH 297

Query: 300 LLTGFDDEESQAIAQMIDQKNEERKEIVQTIFDQAMQMLDQ---TKPVQVLAKENWHPGV 356
           LL   D  E++ +A  IDQ N+ER+++V  + D+A++M++Q    +   V+AK  W+PGV
Sbjct: 298 LLMSEDSFEAEELAAEIDQLNKERQKMVSKMTDEAIEMVEQQGLDQTAIVVAKAGWNPGV 357

Query: 357 LGIVAGRILERTGQPVIVLNI--EDGIAKGSARSVEALDIFQAFDQHRELFIAFGGHSGA 414
           +GIVA ++++R  +P IVL I  E GIAKGSARS+   ++F++  + R++   FGGH  A
Sbjct: 358 VGIVASKLVDRFYRPAIVLGIDEEKGIAKGSARSIRGFNLFESLSECRDILPHFGGHPMA 417

Query: 415 AGMTLEESKVGDLSQVLCDYISKKQLDMSQKKTLTIDSELRFDELSLDTVRDFEKLAPFG 474
           AGMTL+   V DL   L +       +        +D     ++++++++ +   L+PFG
Sbjct: 418 AGMTLKAEDVPDLRSRLNEIADNTLTEEDFIPVQEVDLVCGVEDITVESIAEMNMLSPFG 477

Query: 475 MDNKKPVFLLKDFKVSQARVMGQNGAHLKLKLEQDGQALDLVAFNMGSQLQEFQQAQHLE 534
           M N KP  L+++  +   R +G N  H+K+ +  +   LD V FN G   +       +
Sbjct: 478 MLNPKPHVLVENAVLEDVRKIGANKTHVKMTIRNESSQLDCVGFNKGELQEGIVPGSRIS 537

Query: 535 LAVTLSVNQWNGATTLQLMLEDARVDGIQLFDIRSK------ASSLPHG----------- 577
           +   +S+N+WN     QLM++DA V   QLFD+R K       S+LP
Sbjct: 538 IVGEMSINEWNNRKKPQLMIKDAAVSEWQLFDLRGKRTWEDTVSALPSAKRAIVSFKEDS 597

Query: 578 ------------VPILSQEEQSKE-------VILLTVPDHPQELKQMTQGKQFDAIYFKN 618
                       V ++S ++Q+K        ++LL  P     L ++ +GK  + IYF
Sbjct: 598 TTLLQTEDLRREVHVISSKDQAKAFDLDGAYIVLLDPPPSLDMLARLLEGKAPERIYFIF 657

Query: 619 EIPKNYFISGYGTRDQFASLYKTIYQFPEFDVRYKLKELSSYLHIPDILLIKMIQIFEEL 678
              +++F+S +  RD F   Y  + +   FDV+    EL+ +       +  M ++F +L
Sbjct: 658 LNHEDHFLSTFPARDHFKWYYAFLLKRGAFDVKKHGSELAKHKGWSVETINFMTKVFFDL 717

Query: 679 HFVTITEGIMTVNKEAEKRDISESQIYQELKETVKFQELMALGTPKEIYDFM 730
            FV I  G+++V   A+KRD+++SQ YQ  ++ ++  + +   + +E+ +++
Sbjct: 718 GFVKIENGVLSVVSGAKKRDLTDSQTYQAKQQLMELDQKLNYSSAEELKEWL 769
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2631> which encodes the amino acid sequence <SEQ ID 2632>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16    Transmembrane  220 - 236 ( 220 - 236)
   INTEGRAL    Likelihood = -0.11    Transmembrane  667 - 683 ( 667 - 683)

----- Final Results -----
               bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 474/731 (64%), Positives = 594/731 (80%)

Query: 1   MISAKYSWVLNNQKPDAGFFEASKKEKISEAVASLIYSRGIKTSAELHHFLQTNLENLHD 60
           MI +KYSW + ++KPD GFF+ +K + +++  A LIY RGI+T   L  FL  +L  LHD
Sbjct: 1   MIKSKYSWKIKDKKPDDGFFKLAKTKGLTQTAAQLIYDRGIRTEEALDEFLTADLSQLHD 60

Query: 61  PYLLNDMDKAVNRIRRAIENNETILVYGDYDADGMTSASIMKEALDMMGAEVQVYLPNRF 120
```

```
           PYLL+DM KAV RIR+AIE  E IL+YGDYDADGMTSASI+KE LDMMGAE  VYLPNRF
Sbjct: 61  PYLLHDMAKAVPRIRQAIEEGERILIYGDYDADGMTSASIVKETLDMMGAEPLVYLPNRF 120

Query: 121 TDGYGPNQSVYKYFIEQQDVSLIITVDNGVAGHEAITYAQNQGVDVVVTDHHSMPADLPC 180
           TDGYGPNQSVYKYFIEQ+ VSLIITVDNGVAGHEAI YAQ Q VDV+VTDHHS+P +LP
Sbjct: 121 TDGYGPNQSVYKYFIEQEAVSLIITVDNGVAGHEAIRYAQEQEVDVIVTDHHSLPEELPE 180

Query: 181 AYAIIHPEHPDANYPFPYLAGCGVAFKVACALLETIPTEMLDLVAIGTIADMVSLTDENR 240
           A+AIIHPEHPDA+YPF +LAGCGVAFK+A ALLE++PT+ LDLVAIGTIADMVSLT ENR
Sbjct: 181 AFAIIHPEHPDADYPFKHLAGCGVAFKLATALLESLPTDCLDLVAIGTIADMVSLTGENR 240

Query: 241 IMVKAGLEVMKDSERIGLQELISLSNIDLKTLNEETIGFKIAPQLNALGRLDDPNPAIEL 300
           ++VK GL ++K +ER+GLQEL+SLS IDL+  NE+ IGF+IAPQLNALGRLDDPNPAIEL
Sbjct: 241 VLVKNGLAMLKHTERVGLQELMSLSPIDLEHFNEDAIGFQIAPQLNALGRLDDPNPAIEL 300

Query: 301 LTGFDDEESQAIAQMIDQKNEERKEIVQTIFDQAMQMLDQTKPVQVLAKENWHPGVLGIV 360
           LTGFDD+E+QAIA MI +KNEERK +VQ IFDQAM M+D  KPVQVLA+  WHPGVLGIV
Sbjct: 301 LTGFDDQEAQAIALMIKKKNEERKALVQDIFDQAMAMVDPQKPVQVLAQAGWHPGVLGIV 360

Query: 361 AGRILERTGQPVIVLNIEDGIAKGSARSVEALDIFQAFDQHRELFIAFGGHSGAAGMTLE 420
           AGRI+E  GQ V+VL I++G AKGSARS+EA++IF+A +  RELF AFGGH+GAAGMTL
Sbjct: 361 AGRIMETIGQTVVVLTIDNGFAKGSARSLEAINIFEALNGKRELFTAFGGHAGAAGMTLP 420

Query: 421 ESKVGDLSQVLCDYISKKQLDMSQKKTLTIDSELRFDELSLDTVRDFEKLAPFGMDNKKP 480
              +  LS  LC ++ ++ LD + K TLTID  L  D+LSLD ++  +KLAP+GMD++KP
Sbjct: 421 VDNLEALSDFLCQFVIERGLDQTAKNTLTIDERLSLDDLSLDILKSLDKLAPYGMDHQKP 480

Query: 481 VFLLKDFKVSQARVMGQNGAHLKLKLEQDGQALDLVAFNMGSQLQEFQQAQHLELAVTLS 540
           VF +KD +VSQAR +GQ+ +HLK K+ Q   + D++AF  GSQLQEF+QA  LELAVTLS
Sbjct: 481 VFYVKDIRVSQARTIGQDQSHLKFKVSQGKASFDVLAFGQGSQLQEFRQATGLELAVTLS 540

Query: 541 VNQWNGATTLQLMLEDARVDGIQLFDIRSKASSLPHGVPILSQEEQSKEVILLTVPDHPQ 600
           VN WNG T+LQ ML DARVDG+QL D+R+K + +P G+P + ++  ++ +++  +P+  +
Sbjct: 541 VNHWNGNTSLQFMLVDARVDGVQLLDLRTKTAKVPEGIPTIEEDPNARVILINDIPEDFK 600

Query: 601 ELKQMTQGKQFDAIYFKNEIPKNYFISGYGTRDQFASLYKTIYQFPEFDVRYKLKELSSY 660
             +     K FDAIYFKN++   Y+++G+G+R+QFA LYKTIYQFPEFD+R+KL ELS Y
Sbjct: 601 TWRNQFVHKDFDAIYFKNQMKHPYYLTGFGSREQFAKLYKTIYQFPEFDLRHKLTELSHY 660

Query: 661 LHIPDILLIKMIQIFEELHFVTITEGIMTVNKEAEKRDISESQIYQELKETVKFQELMAL 720
           L+I  +LLIK+IQIFEEL FVTI +G+MTVN +A+KR+ISES IYQ+LKE VKFQE+MAL
Sbjct: 661 LNIEKLLLIKLIQIFEELSFVTIDDGLMTVNPQAQKREISESHIYQDLKELVKFQEIMAL 720

Query: 721 GTPKEIYDFMM 731
            +PKE+YD+++
Sbjct: 721 ASPKEMYDYLV 731
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 868

A DNA sequence (GBSx0920) was identified in *S.agalactiae* <SEQ ID 2633> which encodes the amino acid sequence <SEQ ID 2634>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4114(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 869

A DNA sequence (GBSx0921) was identified in *S.agalactiae* <SEQ ID 2635> which encodes the amino acid sequence <SEQ ID 2636>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.10    Transmembrane   15 -   31 (   14 -   33)

----- Final Results -----
               bacterial membrane --- Certainty=0.3039(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA88584 GB:M18954 fructosyltransferase [Streptococcus mutans]
 Identities = 67/219 (30%), Positives = 106/219 (47%), Gaps = 31/219 (14%)

Query: 1   MRPIVRKKMYKKGKFWVVAGIVT-ILGGSAILGQDVKAEQAEAVTSTISEKTDSSQTISD 59
           M   VRKKMYKKGKFWVVA I T +L G  +    V+A++A + T   SE  + SQ   +
Sbjct: 1   METKVRKKMYKKGKFWVVATITTAMLTGIGL--SSVQADEANS-TQVSSELAERSQVQEN 57

Query: 60  TSKLTLPVNSSEAMKNSAEPLIKTAFATSVSSNPREIAATPVKTFDASSKVVVKASTAEH 119
           T+       SS A +N A    KT    + S+NP   AA  V+  D ++KV+   +  E
Sbjct: 58  TTA------SSSAAENQA----KTEVQETPSTNP---AAATVENTDQTTKVITDNAAVES 104

Query: 120 SANQTN---SNVNQVANDSEVITQQN------STKQLPTVTYSAHVQDIGW----QKSVD 166
            A++T    + V + A  +  + Q N      +TK+    T    + + G     +K
Sbjct: 105 KASKTKDQAATVTKTAASTPEVGQTNEKDKAKATKEADITTPKNTIDEYGLTEQARKIAT 164

Query: 167 NATVSGTVGQEKQVEAIKLSIKAPEGITG-KLSYKTYVK 204
            A ++ +   +KQVEA+       +  TG +++Y+ + K
Sbjct: 165 EAGINLSSLTQKQVEALNKVKLTSDAQTGHQMTYQEFDK 203
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8677> and protein <SEQ ID 8678> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 5
McG: Discrim Score:      9.08
GvH: Signal Score (-7.5): -3.94
     Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 1 value:  -5.10 threshold:  0.0
   INTEGRAL    Likelihood = -5.10    Transmembrane    7 -  23 (   6 -  25)
   PERIPHERAL  Likelihood =  4.03     694
 modified ALOM score:   1.52

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3039(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

31.1/52.1% over 749aa

Streptococcus mutans

EGAD|14681| levansucrase precursor Insert characterized

SP|P11701|SACB_STRMU LEVANSUCRASE PRECURSOR (EC 2.4.1.10) (BETA-D-FRUCTOFURANOSYL TRANSFERASE) (SUCROSE 6-FRUCTOSYL TRANSFERASE). Edit characterized

GP|153636|gb|AAA88584.1||M18954 fructosyltransferase Insert characterized

PIR|B28551|B28551 levansucrase (EC 2.4.1.10) precursor - (strain GS-5) Insert characterized

ORF02172(295 - 1731 of 3138)
EGAD|14681|14686(7 - 756 of 797) levansucrase precursor {Streptococcus mutans} SP|P11701|SACB_STRMU LEVANSUCRASE PRECURSOR (EC 2.4.1.10) (BETA-D-FRUCTOFURANOSYL TRANSFERASE) (SUCROSE 6-FRUCTOSYL TRANSFERASE). GP|153636|gb|AAA88584.1||M18954 fructosyltransferase {Streptococcus mutans} PIR|B28551|B28551 levansucrase (EC 2.4.1.10) precursor - Streptococcus mutans (strain GS-5)
%Match = 2.9
%Identity = 31.1 %Similarity = 52.1
Matches = 83 Mismatches = 115 Conservative Sub.s = 56

```
132       162       192       222       252       282       312       342
LPEHLENQSYQH*PYQH*YQ*RHNHHQYLVQ*ERVQQLIQRAPCL*FQFYVSYXXXN*LXXYR*KKMYKKGKFWVVAGIV
                                                          :     ||||||||||||| |
                                                          METKVRKKMYKKGKFWVVATIT
                                                                  10        20

372       402       432       462       492       522       552       582
TILGGSAILGQDVKAEQAEAVTSTISEKTDSSQTISDTSKLTLPVNSSEAMKNSAEPLIKTAFATSVSSNPREIAATPVK
| :    : |  : : ||:   ||        |  |: :: :    :: |  ::||   ||    : |:||   ||  |:
TAM----LTGIGLSSVQADEANST--------QVSSELAERSQVQENTTASSSAAENQAKTEVQETPSTNP---AAATVE
          30        40                50        60        70        80

612       642       663       693       705       735                 783
TFDASSKVVVKASTAEHSANQTNSN---VNQVANDSEVITQQN------STKQLPTVTYSAHVQDIGW----QKSVDNAT
  | ::||:   :  |  |::|      | : |  :  : | |      :||:    |    : : |     :|    |
NTDQTTKVITDNAAVESKASKTKDQAATVTKTAASTPEVGQTNEKDKAKATKEADITTPKNTIDEYGLTEQARKIATEAG
          100       110       120       130       140       150       160

813       834                                                         882
VSGTVGQEKQVEA---IKLSIKAPEG--------------------~~~~--------------------ITGKLSYKTY
:: :   :|||||   :||:  |  |                                            : |    | :
INLSSLTQKQVEALNKVKLTSDAQTGHQMTYQEFDKIAQTLIAQDE~~~~VGTLDTAYLPGENDGYIDWNVIGGYGLKPH
          180       190       200       210               660       670

912       942       972       1002      1032      1062      1092      1122
VKGQGWQPSVESGQVSGTVGQSRPIEALSINLTDNLQKLYDVYYRVHVQDIGWMAWAKNGAYAGTLGMSKRLEAYEVKFT
  || :||:|
TPGQ-YQPTV----------------------------------------------------------------------

1152      1182      1209      1239      1269      1290      1320      1350
LKGQSVLTPTIPKEERPVLNYQVKV-GQNGWQSNKLEGQMAGTLGESKALDG---VKFTLSTLKYGDILYRTHVQDKGWG
        |: |     :::::|   |:   :  |:    || : :|:   |   | | :|
--------PSTPIHTDDIISFEVSFDGHLVIKPVKVNNDSAGRIDQSRNSGGSLNVAFNVSA------------------
       690       700       710       720       730       740

1641      1671      1701      1731      1761      1791      1821      1851
EI~~~~SYQTYLQKDGWKPTVLEGQLGGSIGLSKSIKAIKLNLGSTALGNIEYRTFLNGSGWQTVVNSGRESNVPNESQQ
                          ||:| :  | |:|     :    ::                    |
--~~~~~~------------------GGNISVKPSQKSINNTKETKKAHHVSTEKKQKKGNSFFAALLALFSAFCVSIGF
                               750       760       770       780       790
```

SEQ ID 8678 (GBS243) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 7; MW 94kDa).

GBS243-His was purified as shown in Figure 208, lane 10.

Example 870

A DNA sequence (GBSx0922) was identified in *S.agalactiae* <SEQ ID 2637> which encodes the amino acid sequence <SEQ ID 2638>. This protein is predicted to be adenine phosphoribosyltransferase (apt). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.86   Transmembrane   61 -  77 (  59 -  77)
   INTEGRAL    Likelihood = -0.64   Transmembrane  137 - 153 ( 137 - 153)

----- Final Results -----
               bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC46040 GB:U86377 adenine phosphoribosyltransferase; Apt
           [Bacillus subtilis]
 Identities = 110/170 (64%), Positives = 135/170 (78%)

Query: 1   MDLNNYIASIENYPQEGITFRDISPLMADGKAYSYAVREIVQYAADKDIDMIVGPEARGF 60
           MDL  Y+  + +YP+EG+ F+DI+ LM  G  Y YA  +IV+YA +K ID++VGPEARGF
Sbjct: 1   MDLKQYVTIVPDYPKEGVQFKDITTLMDKGDVYRYATDQIVEYAKEKQIDLVVGPEARGF 60

Query: 61  IVGCPVAYALGIGFAPVRKPGKLPREVISADYEKEYGLDTLTMHADAIKPGQRVLIVDDL 120
           I+GCPVAYALG+GFAPVRK GKLPREVI  DY  EYG D LT+H DAIKPGQRVLI DDL
Sbjct: 61  IIGCPVAYALGVGFAPVRKEGKLPREVIKVDYGLEYGKDVLTIHKDAIKPGQRVLITDDL 120

Query: 121 LATGGTVKATIEMIEKLGGVVAGCAFLVELDGLNGRKAIEGYDTKVLMNF 170
           LATGGT++ATI+++E+LGGVVAG AFL+EL  L+GR  +E YD   LM +
Sbjct: 121 LATGGTIEATIKLVEELGGVVAGIAFLIELSYLDGRNKLEDYDILTLMKY 170
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2639> which encodes the amino acid sequence <SEQ ID 2640>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty= 0.300(Affirmative) < succ>
               bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
!GB:Z99120 similar to opine catabolism [Bacillus sub...    231  1e-59

>GP:CAB15253 GB:Z99120 similar to opine catabolism [Bacillus subtilis]

 Score =  231 bits (583), Expect = 1e-59
 Identities = 138/363 (38%), Positives = 212/363 (58%), Gaps = 11/363 (3%)

Query: 5   IIGAGIVGSTAAYYLQQSGQKEVTIFDHGQ-GQATKAAAGIISPWFSKRRNKVWYRMARL 63
           I+GAGI+G++ AY+L ++G + VT+ D  + GQAT AAAGI+ PW S+RRN+ WY++A+
Sbjct: 6   IVGAGILGASTAYHLAKTGAR-VTVIDRKEPGQATDAAAGIVCPWLSQRRNQDWYQLAKG 64

Query: 64  GADFYQQLINDLKEDGFATDFYQQNGIYVLKKQEEKLRDLYELALARKVESPIIGELAIK 123
           GA +Y+ LI+ L++DG +   Y++ G   +     KL  + E A  R+ ++P IG++
Sbjct: 65  GARYYKDLIHQLEKDGESDTGYKRVGAISIHTDASKLDKMEERAYKRREDAPEIGDITRL 124

Query: 124 NRKELGNDFKGLIGFDNCLYASGAARVEGAALCETLLKAS---GYPVIRQKVTLKQQG-- 178
```

```
           +  E    F  L      ++ SGAARV G ALC +LL A+   G  VI+   +L  +
Sbjct: 125 SASETKKLFPILADGYESVHISGAARVNGRALCRSLLSAAEKRGATVIKGNASLLFENGT 184

Query: 179 -SGYEIAGHYF--DQVILAAGAWLPDLLRPLGYQVDVRPQKGQLLDYDVHHIISDTYPVV 235
            +G +     F  D VI+ AGAW  ++L+PLG    V  QK Q++ +++    + ++PVV
Sbjct: 185 VTGVQTDTKQFAADAVIVTAGAWANEILKPLGIHFQVSFQKAQIMHFEMTDADTGSWPVV 244

Query: 236 MPEGEIDLIPFNQGKISVGTSHENDKGY-DLEPDWQVLKKLEMQALTYLPLLKEATQKTC 294
           MP  +  ++ F+ G+I  G +HEND G  DL        ++  +AL   P L +A
Sbjct: 245 MPPSDQYILSFDNGRIVAGATHENDAGLDDLRVTAGGQHEVLSKALAVAPGLADAAAVET 304

Query: 295 RVGIRAYTSDYSPFYGQVSGLKNLYTASGLGSSGLTVGPLIGYELAQLLLGHEGLLTPSD 354
           RVG R +T  + P  G V  ++ LY A+GLG+SGLT+GP +G ELA+L+LG +  L  S
Sbjct: 305 RVGFRPFTPGFLPVVGAVPNVQGLYAANGLGASGLTMGPFLGAELAKLVLGKQTELDLSP 364

Query: 355 YSP 357
           Y P
Sbjct: 365 YDP 367
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 150/172 (87%), Positives = 161/172 (93%)

Query: 1   MDLNNYIASIENYPQEGITFRDISPLMADGKAYSYAVREIVQYAADKDIDMIVGPEARGF 60
           MDL NYIASI++YP+ GITFRDISPLMADGKAYSYA+REI QYA DKDIDM+VGPEARGF
Sbjct: 1   MDLTNYIASIKDYPKAGITFRDISPLMADGKAYSYAIREIAQYACDKDIDMVVGPEARGF 60

Query: 61  IVGCPVAYALGIGFAPVRKPGKLPREVISADYEKEYGLDTLTMHADAIKPGQRVLIVDDL 120
           I+GCPVA  LGIGFAPVRKPGKLPR+V+SADYEKEYGLDTLTMHADAIKPGQRVLIVDDL
Sbjct: 61  IIGCPVAVELGIGFAPVRKPGKLPRDVVSADYEKEYGLDTLTMHADAIKPGQRVLIVDDL 120

Query: 121 LATGGTVKATIEMIEKLGGVVAGCAFLVELDGLNGRKAIEGYDTKVLMNFPG 172
           LATGGTVKATIEMIEKLGG+VAGCAFL+EL+GLNGR AI  YD KVLM FPG
Sbjct: 121 LATGGTVKATIEMIEKLGGIVAGCAFLIELEGLNGRHAIRNYDYKVLMQFPG 172
```

SEQ ID 2638 (GBS419) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 6; MW 22.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 4; MW 47.5kDa).

GBS419-GST was purified as shown in Figure 219, lane 6-8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 871

A DNA sequence (GBSx0923) was identified in *S.agalactiae* <SEQ ID 2641> which encodes the amino acid sequence <SEQ ID 2642>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0847(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA11244 GB:D78182 ORF2 [Streptococcus mutans]
 Identities = 140/225 (62%), Positives = 178/225 (78%)

Query: 1   MTYLEQYQSGQLTLPSALFFHFKSIFKTADDFLVWQFFYLQNTTNLSDLTPSRIATSLDK 60
           M++L+ Y+SG L LPSAL FH+K IF  ADDFLVWQFFY QNTT + D+  S+IAT++ K
```

```
Sbjct: 1    MSFLQHYKSGNLVLPSALLFHYKDIFSNADDFLVWQFFYFQNTTKMEDIATSQIATAIGK 60

Query: 61   TVADINRSISNLTSQGLLDVKTIELNHEIEIIFDTSPVFAKLDKLFEEDNQVIIDNKTSD 120
            TV ++NRS+SNL SQ LLD+KTIEL+ E E++FD +    KLD L    ++  + +
Sbjct: 61   TVPEVNRSVSNLISQELLDMKTIELDGESEVLFDATLALKKLDDLLTAADETTVSSSKGT 120

Query: 121  SNRLKDLVGDFERELGRLLSPFELEDLQKTLQEDQTDPDIVRAALREAVFNGKTSWNYIN 180
            SN LKDLV DFERELGR+LSPFELEDLQKT+ +D+TDPD+VR+ALREAVFNGKT+WNYI
Sbjct: 121  SNALKDLVEDFERELGRMLSPFELEDLQKTVSDDKTDPDLVRSALREAVFNGKTNWNYIQ 180

Query: 181  AILRNWRREGLTTLRQIEERKQAREDNQMKDLAISDDFKNAMNLW 225
            AILRNWRREG++TLRQ+EER++ RE      ++ +SDDF +AMNLW
Sbjct: 181  AILRNWRREGISTLRQVEERRKEREQANPANVTVSDDFLSAMNLW 225
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2643> which encodes the amino acid sequence <SEQ ID 2644>. Analysis of this protein sequence reveals the following:

```
      Possible site: 57

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAA11244 GB:D78182 ORF2 [Streptococcus mutans]
 Identities = 154/228 (67%), Positives = 188/228 (81%), Gaps = 1/228 (0%)

Query: 1    MSFLEHYKSGNLVIPSALLFHYKDLFKSSDDFLVWQFFYLQNTTKRDDLAPSQIAHALGK 60
            MSFL+HYKSGNLV+PSALLFHYKD+F ++DDFLVWQFFY QNTTK +D+A SQIA A+GK
Sbjct: 1    MSFLQHYKSGNLVLPSALLFHYKDIFSNADDFLVWQFFYFQNTTKMEDIATSQIATAIGK 60

Query: 61   SVADINKIISSLTNQGLLDMRTIELTGEIEIIFDASPVLAKLDQLFVSQTATEIDKQE-T 119
            +V ++N+ +S+L +Q LLDM+TIEL GE E++FDA+  L KLD L  +   T +   + T
Sbjct: 61   TVPEVNRSVSNLISQELLDMKTIELDGESEVLFDATLALKKLDDLLTAADETTVSSSKGT 120

Query: 120  PNHFKRLVDEFERELGRFLSPFELEDLEKTLRDDKTDPDLIREALKEAVFNGKTNWKYIQ 179
             N  K LV++FERELGR LSPFELEDL+KT+ DDKTDPDL+R AL+EAVFNGKTNW YIQ
Sbjct: 121  SNALKDLVEDFERELGRMLSPFELEDLQKTVSDDKTDPDLVRSALREAVFNGKTNWNYIQ 180

Query: 180  AILRNWRKEGIVNLRQVEERRRVREGEDLSQVTISEDFLSAMNLWSDS 227
            AILRNWR+EGI  LRQVEERR+ RE  + + VT+S+DFLSAMNLWSDS
Sbjct: 181  AILRNWRREGISTLRQVEERRKEREQANPANVTVSDDFLSAMNLWSDS 228
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 144/225 (64%), Positives = 179/225 (79%), Gaps = 1/225 (0%)

Query: 1    MTYLEQYQSGQLTLPSALFFHFKSIFKTADDFLVWQFFYLQNTTNLSDLTPSRIATSLDK 60
            M++LE Y+SG L +PSAL FH+K +FK++DDFLVWQFFYLQNTT   DL PS+IA +L K
Sbjct: 1    MSFLEHYKSGNLVIPSALLFHYKDLFKSSDDFLVWQFFYLQNTTKRDDLAPSQIAHALGK 60

Query: 61   TVADINRSISNLTSQGLLDVKTIELNHEIEIIFDTSPVFAKLDKLFEEDNQVIIDNKTSD 120
            +VADIN+ IS+LT+QGLLD++TIEL  EIEIIFD SPV AKLD+LF       ID K
Sbjct: 61   SVADINKIISSLTNQGLLDMRTIELTGEIEIIFDASPVLAKLDQLFVSQTATEID-KQET 119

Query: 121  SNRLKDLVGDFERELGRLLSPFELEDLQKTLQEDQTDPDIVRAALREAVFNGKTSWNYIN 180
             N  K LV +FERELGR LSPFELEDL+KTL++D+TDPD++R AL+EAVFNGKT+W YI
Sbjct: 120  PNHFKRLVDEFERELGRFLSPFELEDLEKTLRDDKTDPDLIREALKEAVFNGKTNWKYIQ 179

Query: 181  AILRNWRREGLTTLRQIEERKQAREDNQMKDLAISDDFKNAMNLW 225
            AILRNWR+EG+  LRQ+EER++ RE   +   + IS+DF +AMNLW
Sbjct: 180  AILRNWRKEGIVNLRQVEERRRVREGEDLSQVTISEDFLSAMNLW 224
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 872

A DNA sequence (GBSx0924) was identified in *S.agalactiae* <SEQ ID 2645> which encodes the amino acid sequence <SEQ ID 2646>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1617(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA11245 GB:D78182 ORF3 [Streptococcus mutans]
 Identities = 134/226 (59%), Positives = 170/226 (74%)

Query: 2   DLQLSKRLQKVANYVPKGARLLDVGSDHAYLPIFLLQMGYCDFAIAGEVVNGPYQSALKN 61
           ++ LS RLQ+VA++VPKGARLLDVGSDHAYLPI+LL+ G  DFA+AGE++ GPY+SA+ N
Sbjct: 7   EVSLSHRLQEVASFVPKGARLLDVGSDHAYLPIYLLEQGLIDFAVAGEIIKGPYESAVAN 66

Query: 62  VSEHGLTSKIDVRLANGLSAFEEADNIDTITICGMGGRLIADILNNDIDKLQHVKTLVLQ 121
           V+E GL+ +I VRLA+GL+A  + D+ID ITICGMGGRLIADIL    DKL  VK L+LQ
Sbjct: 67  VNESGLSGQIAVRLADGLAALNDNDDIDLITICGMGGRLIADILAAGSDKLNSVKQLILQ 126

Query: 122 PNNREDDLRKWLAANDFEIVAEDILTENDKRYEILVVKHGHMNLTAKELRFGPFLLSNNT 181
           PNN EDDLR WL ANDF I AE ++ +  K YEILVV+ G + L+ K+LRFGPFL    +
Sbjct: 127 PNNCEDDLRSWLVANDFMIKAEKMVKDRHKYYEILVVEKGKITLSDKDLRFGPFLRQERS 186

Query: 182 TVFKEKWQNELNKLTFALNSIPNSKMEERAILEDKIQDIKEVLDES 227
           ++FKE+W+ EL KL  AL  +P  K  +   L  KI+ I+EVL ES
Sbjct: 187 SIFKERWRKELAKLELALTRVPAKKKADNMFLSTKIEQIREVLYES 232
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2647> which encodes the amino acid sequence <SEQ ID 2648>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0803(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 145/224 (64%), Positives = 173/224 (76%)

Query: 1   MDLQLSKRLQKVANYVPKGARLLDVGSDHAYLPIFLLQMGYCDFAIAGEVVNGPYQSALK 60
           MD QLS RL +VA YVPKG +LLDVGSDHAYLPIFL++      AIAGEVV GPY+SALK
Sbjct: 1   MDSQLSNRLAQVAAYVPKGVKLLDVGSDHAYLPIFLVETNQISAAIAGEVVRGPYESALK 60

Query: 61  NVSEHGLTSKIDVRLANGLSAFEEADNIDTITICGMGGRLIADILNNDIDKLQHVKTLVL 120
           NV++ GL   I VRLANGL+AFEEAD++  ITICGMGGRLIADIL    +KLQ ++ LVL
Sbjct: 61  NVTQSGLAEHIQVRLANGLAAFEEADDVTAITICGMGGRLIADILEAGKEKLQGIERLVL 120

Query: 121 QPNNREDDLRKWLAANDFEIVAEDILTENDKRYEILVVKHGHMNLTAKELRFGPFLLSNN 180
           QPNNREDDLR WL+ N F+IVAE I+ ENDK YEI+V +HG   L+A ELRFGP+L
Sbjct: 121 QPNNREDDLRAWLSVNAFKIVAETIMAENDKYYEIIVAEHGEKALSATELRFGPYLSQEK 180

Query: 181 TTVFKEKWQNELNKLTFALNSIPNSKMEERAILEDKIQDIKEVL 224
```

```
           + VFKEKWQ E++KL +AL+ IP  K +ER +L  KIQ IKEV+
Sbjct: 181 SVVFKEKWQREMDKLAYALSCIPEEKTQERQLLLTKIQQIKEVI 224
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 873

A DNA sequence (GBSx0925) was identified in *S.agalactiae* <SEQ ID 2649> which encodes the amino acid sequence <SEQ ID 2650>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3245(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9893> which encodes amino acid sequence <SEQ ID 9894> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA11246 GB:D78182 ORF4 [Streptococcus mutans]
 Identities = 187/262 (71%), Positives = 224/262 (85%)

Query: 2   MKARELIDVYETYCPQELSMEGDISGLQIGSLDKEIKTVMVALDVRETTVAEAIERQVDL 61
           MKA ++I  YE YCPQ+LS+EGDISGLQIG+LDKEIK +M+ALDVRETTVAEAIE++VDL
Sbjct: 1   MKASQIIKRYEAYCPQDLSLEGDISGLQIGTLDKEIKRLMIALDVRETTVAEAIEKKVDL 60

Query: 62  LIVKHAPIFRPLKDLVATPQNKIYIDLLKSDIAVYVSHTNIDIVPNGLNDWFCELLDIQY 121
           LIVKHAPIFRPLK+LV T QN IY +L+K DIAVYVSHTNIDIVP+GLNDWFC+LLDI+
Sbjct: 61  LIVKHAPIFRPLKNLVETAQNHIYFNLIKHDIAVYVSHTNIDIVPDGLNDWFCDLLDIKN 120

Query: 122 PDILSETSNGYGIGRIGDIRPQSFEFFAWKIKDVFGLDSVRLVSYDKSNPEIQRVAICGG 181
             ILS + + YGIGR+GDI P SFE  A K+K +F LDSVRLVSY ++NP I R+AICGG
Sbjct: 121 RRILSPSKDDYGIGRVGDISPLSFEDLAKKVKKIFNLDSVRLVSYGENNPLISRIAICGG 180

Query: 182 SGQSFYKEAIAKGADVFVTGDIYYHTAQEMITNGLLAIDPGHHIEVLFVSKIATMIEQWK 241
           SGQSFY+EA+ KGA V++TGDIYYHTAQEM+TNGLLA+DPGHHIEVLFV K+A   + W
Sbjct: 181 SGQSFYQEALTKGAQVYITGDIYYHTAQEMLTNGLLALDPGHHIEVLFVRKLAEKFQTWS 240

Query: 242 LEKGWDISVLESKAPTNPFYHM 263
            ++ WDI++LES+  TNPFYH+
Sbjct: 241 CQENWDITILESQVNTNPFYHL 262
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2651> which encodes the amino acid sequence <SEQ ID 2652>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1804(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 169/262 (64%), Positives = 214/262 (81%)

Query: 2   MKARELIDVYETYCPQELSMEGDISGLQIGSLDKEIKTVMVALDVRETTVAEAIERQVDL 61
           MKA+ LID YE +CP +LSMEGD+ GLQ+GSLDK+I+ VM+ LD+RE+TVAEAI+ +VDL
```

```
Sbjct: 3   MKAKTLIDAYEAFCPLDLSMEGDVKGLQMGSLDKDIRKVMITLDIRESTVAEAIKNEVDL 62

Query: 62  LIVKHAPIFRPLKDLVATPQNKIYIDLLKSDIAVYVSHTNIDIVPNGLNDWFCELLDIQY 121
           +I KHAPIF+PLKDLV++PQ  I +DL+K DI+VYVSHTNIDIVP GLNDWFC+LL+I+
Sbjct: 63  IITKHAPIFKPLKDLVSSPQRDILLDLVKHDISVYVSHTNIDIVPGGLNDWFCDLLEIKE 122

Query: 122 PDILSETSNGYGIGRIGDIRPQSFEFFAWKIKDVFGLDSVRLVSYDKSNPEIQRVAICGG 181
              LSET  G+GIGRIG ++ Q+ E  A K+K VF LD+VRL+ YDK NP I ++AICGG
Sbjct: 123 ATYLSETKEGFGIGRIGTVKEQALEELASKVKRVFDLDTVRLIRYDKENPLISKIAICGG 182

Query: 182 SGQSFYKEAIAKGADVFVTGDIYYHTAQEMITNGLLAIDPGHHIEVLFVSKIATMIEQWK 241
           SG  FY++A+ KGADV++TGDIYYHTAQEM+T GL A+DPGHHIEVLF  K+   ++ WK
Sbjct: 183 SGGEFYQDAVQKGADVYITGDIYYHTAQEMLTEGLFAVDPGHHIEVLFTEKLKEKLQGWK 242

Query: 242 LEKGWDISVLESKAPTNPFYHM 263
            E GWD+S++ SKA TNPF H+
Sbjct: 243 EENGWDVSIISSKASTNPFSHL 264
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 874

A DNA sequence (GBSx0926) was identified in *S.agalactiae* <SEQ ID 2653> which encodes the amino acid sequence <SEQ ID 2654>. This protein is predicted to be (). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB15253 GB:Z99120 similar to opine catabolism [Bacillus subtilis]
 Identities = 148/368 (40%), Positives = 211/368 (57%), Gaps = 13/368 (3%)

Query: 1   MKKIAIIGAGAVGATLAYYLSKEKDIQVTVFDYGV-GQATKAAAGIISPWFSKRRNKAWY 59
           MK   I+GAG +GA+ AY+L+K    +VTV D    GQAT AAAGI+ PW S+RRN+ WY
Sbjct: 1   MKSYIIVGAGILGASTAYHLAKT-GARVTVIDRKEPGQATDAAAGIVCPWLSQRRNQDWY 59

Query: 60  RMARLGADFYSKLVTDLQKDGFETKFYQQTGVFLLKKDESQLESLFALADKRRLESPLIG 119
           ++A+ GA +Y  L+  L+KDG     Y++ G   +  D S+L+ +   A KRR ++P IG
Sbjct: 60  QLAKGGARYYKDLIHQLEKDGESDTGYKRVGAISIHTDASKLDKMEERAYKRREDAPEIG 119

Query: 120 DLQILNKSEANTHFPEL-DGYEQLLYASGGARVEGADLTRILLEAS---GVNVIKDEVHF 175
           D+  L+ SE    FP L DGYE  ++ SG ARV G  L R LL A+   G  VIK
Sbjct: 120 DITRLSASETKKLFPILADGYES-VHISGAARVNGRALCRSLLSAAEKRGATVIKGNASL 178

Query: 176 -----TITDNGFRVQGIDFDKLVLASGAWLAKILDEHNYQVDVRPQKGQLRDYYFSNINT 230
                T+T      +    D +++ +GAW  +IL        V  QK Q+  +  ++ +T
Sbjct: 179 LFENGTVTGVQTDTKQFAADAVIVTAGAWANEILKPLGIHFQVSFQKAQIMHFEMTDADT 238

Query: 231 GKYPVVMPEGELDIIPFDNGKVSVGASHENDMAF-DLNIDFKVLDKFEEQAIGYFPQLKK 289
           G +PVVMP  +  I+ FDNG++  GA+HEND    DL +      +   +A+   P L
Sbjct: 239 GSWPVVMPPSDQYILSFDNGRIVAGATHENDAGLDDLRVTAGGQHEVLSKALAVAPGLAD 298

Query: 290 ADTTSERVGIRAYTSDFSPFFGPVPCMEGAYAASGLGSTGLTVGPLIGYELCQLILNKEN 349
           A     RVG R +T  F P  G VP ++G YAA+GLG++GLT+GP +G EL +L+L K+
Sbjct: 299 AAAVETRVGFRPFTPGFLPVVGAVPNVQGLYAANGLGASGLTMGPFLGAELAKLVLGKQT 358

Query: 350 QLNLEDYD 357
```

```
           +L+L  YD
Sbjct: 359 ELDLSPYD 366
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2655> which encodes the amino acid sequence <SEQ ID 2656>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 211/360 (58%), Positives = 262/360 (72%)

Query: 3   KIAIIGAGAVGATLAYYLSKEKDIQVTVFDYGVGQATKAAAGIISPWFSKRRNKAWYRMA 62
            KIAIIGAG VG+T AYYL +    +VT+FD+G GQATKAAAGIISPWFSKRRNK WYRMA
Sbjct: 2   KIAIIGAGIVGSTAAYYLQQSGQKEVTIFDHGQGQATKAAAGIISPWFSKRRNKVWYRMA 61

Query: 63  RLGADFYSKLVTDLQKDGFETKFYQQTGVFLLKKDESQLESLFALADKRRLESPLIGDLQ 122
            RLGADFY +L+ DL++DGF T FYQQ G+++LKK E +L  L+ LA  R++ESP+IG+L
Sbjct: 62  RLGADFYQQLINDLKEDGFATDFYQQNGIYVLKKQEEKLRDLYELALARKVESPIIGELA 121

Query: 123 ILNKSEANTHFPELDGYEQLLYASGGARVEGADLTRILLEASGVNVIKDEVHFTITDNGF 182
            I N+ E    F  L G++  LYASG ARVEGA L   LL+ASG  VI+ +V       +G+
Sbjct: 122 IKNRKELGNDFKGLIGFDNCLYASGAARVEGAALCETLLKASGYPVIRQKVTLKQQGSGY 181

Query: 183 RVQGIDFDKLVLASGAWLAKILDEHNYQVDVRPQKGQLRDYYFSNINTGKYPVVMPEGEL 242
             + G  FD+++LA+GAWL  +L     YQVDVRPQKGQL DY   +I +  YPVVMPEGE+
Sbjct: 182 EIAGHYFDQVILAAGAWLPDLLRPLGYQVDVRPQKGQLLDYDVHHIISDTYPVVMPEGEI 241

Query: 243 DIIPFDNGKVSVGASHENDMAFDLNIDFKVLDKFEEQAIGYFPQLKKADTTSERVGIRAY 302
            D+IPF+ GK+SVG SHEND  +DL  D++VL K E QA+ Y P LK+A   + RVGIRAY
Sbjct: 242 DLIPFNQGKISVGTSHENDKGYDLEPDWQVLKKLEMQALTYLPLLKEATQKTCRVGIRAY 301

Query: 303 TSDFSPFFGPVPCMEGAYAASGLGSTGLTVGPLIGYELCQLIILNKENQLNLEDYDITKYV 362
            TSD+SPF+G V  ++  Y ASGLGS+GLTVGPLIGYEL QL+L  E  L   DY    Y+
Sbjct: 302 TSDYSPFYGQVSGLKNLYTASGLGSSGLTVGPLIGYELAQLLLGHEGLLTPSDYSPEPYL 361
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8679> and protein <SEQ ID 8680> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 2
McG: Discrim Score:      4.44
GvH: Signal Score (-7.5): 0.81
     Possible site: 41
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:    7.32 threshold:  0.0
   PERIPHERAL  Likelihood =  7.32      153
 modified ALOM score:  -1.96

*** Reasoning Step: 3

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
45.2/62.7% over 163aa
                                                                   Bacillus subtilis
  EGAD|109026| hypothetical protein Insert characterized
   SP|032159|YURR_BACSU HYPOTHETICAL 39.4 KDA OXIDOREDUCTASE IN HOM-MRGA INTERGENIC REGION.
Insert characterized
  GP|2635760|emb|CAB15253.1||Z99120 similar to opine catabolism Insert characterized
  PIR|A70019|A70019 opine catabolism homolog yurR - Insert characterized
```

```
ORF02167(301 - 792 of 1161)
EGAD|109026|BS3258(1  -  164  of  372)  hypothetical  protein  {Bacillus  subtilis}
SP|032159|YURR_BACSU HYPOTHETICAL 39.4 KDA OXIDOREDUCTASE IN HOM-MRGA INTERGENIC REGION.
GP|2635760|emb|CAB15253.1||Z99120  similar  to  opine  catabolism  {Bacillus  subtilis}
PIR|A70019|A70019 opine catabolism homolog yurR - Bacillus subtilis
%Match = 16.6
%Identity = 45.2  %Similarity = 62.7
Matches = 75  Mismatches = 58  Conservative Sub.s = 29

228       258       288       318       348       378                 435
SYYD*AVET*KRLGYFSFRE*SSNKSLLPYVGAIMKKIAIIGAGAVGATLAYYLSKEKDIQVTVFDYGV-GQATKAAAGI
                                  ||   |:||| :||: ||:|:|    :||| |    |||| |||||
                                  MKSYIIVGAGILGASTAYHLAKT-GARVTVIDRKEPGQATDAAAGI
                                          10        20         30        40

465       495       525       555       585       615       645       675
ISPWFSKRRNKAWYRMARLGADFYSKLVTDLQKDGFETKFYQQTGVFLLKKDESQLESLFALADKRRLESPLIGDLQILN
: ||:|:|||: ||::|: || :|  |:  |:|||     |:: |   :  | |:|: :   | ||| ::| |||:  |:
VCPWLSQRRNQDWYQLAKGGARYYKDLIHQLEKDGESDTGYKRVGAISIHTDASKLDKMEERAYKRREDAPEIGDITRLS
             60        70        80        90       100       110       120

705       732       762       792       822       852       882       912
KSEANTHFPEL-DGYEQLLYASGGARVEGADLTRILXEASGVNVIKDESHFTITDKWLSCSRN*F**TCLASGAPAS*IL
 ||    || | ||||  :: || ||| |  | | |  |:                           |   :
ASETKKLFPILADGYE-SVHISGAARVNGRALCRSLLSAAEKRGATVIKGNASLLFENGTVTGVQTDTKQFAADAVIVTA
            140         150      160       170       180       190       200
```

SEQ ID 8680 (GBS290) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 6; MW 22kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 4; MW 47kDa).

GBS290-GST was purified as shown in Figure 226, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 875

A DNA sequence (GBSx0927) was identified in *S.agalactiae* <SEQ ID 2657> which encodes the amino acid sequence <SEQ ID 2658>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.18    Transmembrane   38 -  54 (  36 -  54)

----- Final Results -----
               bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD19913 GB:AF105113 glucose-1-phosphate thymidylyl transferase
            [Streptococcus pneumoniae]
 Identities = 262/289 (90%), Positives = 276/289 (94%)
```

```
Query: 1   MKGIILAGGSGTRLYPLTRAASKQLMPIYDKPMIYYPLSVLMLAGIKEILIISTPQDLPR 60
           MKGIILAGGSGTRLYPLTRAASKQLMP+YDKPMIYYPLS LMLAGIK+ILIISTPQDLPR
Sbjct: 1   MKGIILAGGSGTRLYPLTRAASKQLMPVYDKPMIYYPLSTLMLAGIKDILIISTPQDLPR 60

Query: 61  FEDMLGDGSELGISLSYAEQPSPDGLAQAFIIGEDFIGDDHVALVLGDNIYHGPGLSAML 120
           F+D+L DGSE GI LSYAEQPSPDGLAQAF+IGE+FIGDD VAL+LGDNIYHGPGLS ML
Sbjct: 61  FKDLLLDGSEFGIKLSYAEQPSPDGLAQAFLIGEEFIGDDSVALILGDNIYHGPGLSTML 120

Query: 121 QRAASKESGATVFGYQVKDPERFGVVEFDTDMNAISIEEKPAQPKSNYAVTGLYFYDNDV 180
           Q+AA KE GATVFGYQVKDPERFGVVEFDTDMNAISIEEKP  P+SNYAVTGLYFYDNDV
Sbjct: 121 QKAAKKEKGATVFGYQVKDPERFGVVEFDTDMNAISIEEKPEYPRSNYAVTGLYFYDNDV 180

Query: 181 VEIAKNIKPSPRGELEITDVNKAYLDRGDLSVELMGRGFAWLDTGTHESLLEAAQYIETV 240
           VEIAK IKPS RGELEITDVNKAYL+RGDLSVELMGRGFAWLDTGTHESLLEA+QYIETV
Sbjct: 181 VEIAKQIKPSARGELEITDVNKAYLNRGDLSVELMGRGFAWLDTGTHESLLEASQYIETV 240

Query: 241 QRMQNVQVANLEEIAYRMGYITREQVLELAQPLKKNEYGQYLLRLIGEA 289
           QRMQNVQVANLEEI+YRMGYI+RE VLELAQPLKKNEYG+YLLRLIGEA
Sbjct: 241 QRMQNVQVANLEEISYRMGYISREDVLELAQPLKKNEYGRYLLRLIGEA 289
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2659> which encodes the amino acid sequence <SEQ ID 2660>. Analysis of this protein sequence reveals the following:

```
     Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1585(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 207-209
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC69538 GB:AF057294 Cps23fO [Streptococcus pneumoniae]
 Identities = 263/289 (91%), Positives = 276/289 (95%)

Query: 1   MKGIILAGGSGTRLYPLTRAASKQLMPIYDKPMIYYPLSTLMLAGIKDVLIISTPQDLPR 60
           MKGIILAGGSGTRLYPLTRAASKQLMP+YDKPMIYYPLSTLMLAGI+D+LIISTPQDLPR
Sbjct: 1   MKGIILAGGSGTRLYPLTRAASKQLMPVYDKPMIYYPLSTLMLAGIRDILIISTPQDLPR 60

Query: 61  FEELLGDGSEFGISLSYKEQPSPDGLAQAFIIGEEFIGDDRVALILGDNIYHGNGLTKML 120
           F+ELL DGSEFGI LSY EQPSPDGLAQAFIIGEEFIGDD VALILGDNIYHG GL+ ML
Sbjct: 61  FKELLQDGSEFGIKLSYAEQPSPDGLAQAFIIGEEFIGDDSVALILGDNIYHGPGLSTML 120

Query: 121 QKAAAKEKGATVFGYQVKDPERFGVVEFDENMNAISIEEKPEVPKSHFAVTGLYFYDNDV 180
           QKAA KEKGATVFGY VKDPERFGVVEFDENMNAISIEEKPE P+S++AVTGLYFYDNDV
Sbjct: 121 QKAAKKEKGATVFGYHVKDPERFGVVEFDENMNAISIEEKPEYPRSNYAVTGLYFYDNDV 180

Query: 181 VEIAKNIKPSARGELEITDVNKAYLERGDLSVELMGRGFAWLDTGTHESLLEAAQYIETV 240
           VEIAK+IKPS RGELEITDVNKAYL+RGDLSVELMGRGFAWLDTGTHESLLEA+QYIETV
Sbjct: 181 VEIAKSIKPSPRGELEITDVNKAYLDRGDLSVELMGRGFAWLDTGTHESLLEASQYIETV 240

Query: 241 QRLQNAQVANLEEIAYRMGYISKEDVHKLAQSLKKNEYGQYLLRLIGEA 289
           QR+QN QVANLEEIAYRMGYIS+EDV  LAQSLKKNEYGQYLLRLIGEA
Sbjct: 241 QRMQNVQVANLEEIAYRMGYISREDVLALAQSLKKNEYGQYLLRLIGEA 289
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 257/289 (88%), Positives = 274/289 (93%)

Query: 1   MKGIILAGGSGTRLYPLTRAASKQLMPIYDKPMIYYPLSVLMLAGIKEILIISTPQDLPR 60
           MKGIILAGGSGTRLYPLTRAASKQLMPIYDKPMIYYPLS LMLAGIK++LIISTPQDLPR
Sbjct: 1   MKGIILAGGSGTRLYPLTRAASKQLMPIYDKPMIYYPLSTLMLAGIKDVLIISTPQDLPR 60

Query: 61  FEDMLGDGSELGISLSYAEQPSPDGLAQAFIIGEDFIGDDHVALVLGDNIYHGPGLSAML 120
```

```
            FE++LGDGSE GISLSY EQPSPDGLAQAFIIGE+FIGDD VAL+LGDNIYHG GL+ ML
Sbjct: 61   FEELLGDGSEFGISLSYKEQPSPDGLAQAFIIGEEFIGDDRVALILGDNIYHGNGLTKML 120

Query: 121 QRAASKESGATVFGYQVKDPERFGVVEFDTDMNAISIEEKPAQPKSNYAVTGLYFYDNDV 180
            Q+AA+KE GATVFGYQVKDPERFGVVEFD +MNAISIEEKP  PKS++AVTGLYFYDNDV
Sbjct: 121 QKAAAKEKGATVFGYQVKDPERFGVVEFDENMNAISIEEKPEVPKSHFAVTGLYFYDNDV 180

Query: 181 VEIAKNIKPSPRGELEITDVNKAYLDRGDLSVELMGRGFAWLDTGTHESLLEAAQYIETV 240
            VEIAKNIKPS RGELEITDVNKAYL+RGDLSVELMGRGFAWLDTGTHESLLEAAQYIETV
Sbjct: 181 VEIAKNIKPSARGELEITDVNKAYLERGDLSVELMGRGFAWLDTGTHESLLEAAQYIETV 240

Query: 241 QRMQNVQVANLEEIAYRMGYITREQVLELAQPLKKNEYGQYLLRLIGEA 289
            QR+QN QVANLEEIAYRMGYI++E V +LAQ LKKNEYGQYLLRLIGEA
Sbjct: 241 QRLQNAQVANLEEIAYRMGYISKEDVHKLAQSLKKNEYGQYLLRLIGEA 289
```

There is also homology to SEQ ID 858.

SEQ ID 2658 (GBS296) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 5; MW 35.4kDa).

GBS296-His was purified as shown in Figure 203, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 876

A DNA sequence (GBSx0929) was identified in *S.agalactiae* <SEQ ID 2661> which encodes the amino acid sequence <SEQ ID 2662>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2635(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 877

A DNA sequence (GBSx0930) was identified in *S.agalactiae* <SEQ ID 2663> which encodes the amino acid sequence <SEQ ID 2664>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1868(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2665> which encodes the amino acid sequence <SEQ ID 2666>. Analysis of this protein sequence reveals the following:

```
     Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2818(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 29-31
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC69539 GB:AF057294 Cps23fP [Streptococcus pneumoniae]
 Identities = 168/197 (85%), Positives = 183/197 (92%)

Query: 1   MTETFFDKPLACREIKEIPGLLEFDIPVRGDNRGWFKENFQKEKMLPIGFPERFFEEGKL 60
           MT+ FF K LA R+++ IPG+LEFDIPV GDNRGWFKENFQKEKMLP+GFPE FF EGKL
Sbjct: 1   MTDNFFGKTLAARKVEAIPGMLEFDIPVHGDNRGWFKENFQKEKMLPLGFPESFFAEGKL 60

Query: 61  QNNVSFSRQHVLRGLHAEPWDKYISVADDGKVLGAWVDLREGETFGNVYQTVIDASKGMF 120
           QNNVSFSR++VLRGLHAEPWDKYISVAD GKVLG+WVDLREGETFGN YQTVIDASKG+F
Sbjct: 61  QNNVSFSRKNVLRGLHAEPWDKYISVADGGKVLGSWVDLREGETFGNTYQTVIDASKGIF 120

Query: 121 VPRGVANGFQVLSETVSYSYLVNDYWALDLKPKYAFVNYADPSLGITWENLAAAEVSEAD 180
           VPRGVANGFQVLS+TVSYSYLVNDYWAL+LKPKYAFVNYADPSLGI WEN+A AEVSEAD
Sbjct: 121 VPRGVANGFQVLSDTVSYSYLVNDYWALELKPKYAFVNYADPSLGIEWENIAEAEVSEAD 180

Query: 181 KNHPLLSDVKPLKPKDL 197
           K+HPLL DVKPLK +DL
Sbjct: 181 KHHPLLKDVKPLKKEDL 197
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 157/197 (79%), Positives = 180/197 (90%)

Query: 1   MTEQFFDKELTCRPIEAIPGLLEFDIPVRGDNRGWFKENFQKEKMIPLGFPESFFEADKL 60
           MTE FFDK L CR I+ IPGLLEFDIPVRGDNRGWFKENFQKEKM+P+GFPE FFE  KL
Sbjct: 1   MTETFFDKPLACREIKEIPGLLEFDIPVRGDNRGWFKENFQKEKMLPIGFPERFFEEGKL 60

Query: 61  QNNISFNKKNTLRGLHAEPWDKYVSIADEGRVIGTWVDLREGDSFGNVYQTIIDASKGIF 120
           QNN+SF++++ LRGLHAEPWDKY+S+AD+G+V+G WVDLREG++FGNVYQT+IDASKG+F
Sbjct: 61  QNNVSFSRQHVLRGLHAEPWDKYISVADDGKVLGAWVDLREGETFGNVYQTVIDASKGMF 120

Query: 121 VPRGVANGFQVLSDKAAYTYLVNDYWALELKPKYAFVNYADPNLGIQWENLEEAEVSEAD 180
           VPRGVANGFQVLS+  +Y+YLVNDYWAL+LKPKYAFVNYADP+LGI WENL  AEVSEAD
Sbjct: 121 VPRGVANGFQVLSETVSYSYLVNDYWALDLKPKYAFVNYADPSLGITWENLAAAEVSEAD 180

Query: 181 KNHPLLKDVKPLKKEDL 197
           KNHPLL DVKPLK +DL
Sbjct: 181 KNHPLLSDVKPLKPKDL 197
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 878

A DNA sequence (GBSx0931) was identified in *S.agalactiae* <SEQ ID 2667> which encodes the amino acid sequence <SEQ ID 2668>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3019(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 879

A DNA sequence (GBSx0932) was identified in *S.agalactiae* <SEQ ID 2669> which encodes the amino acid sequence <SEQ ID 2670>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 880

A DNA sequence (GBSx0933) was identified in *S.agalactiae* <SEQ ID 2671> which encodes the amino acid sequence <SEQ ID 2672>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0957(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9367> which encodes amino acid sequence <SEQ ID 9368> was also identified.

The protein is similar to the dTDP-glucose-4,6-dehydratase from *S.mutans:*

```
>GP:BAA11249 GB:D78182 dTDP-glucose-4,6-dehydratase [Streptococcus mutans]
 Identities = 290/310 (93%), Positives = 304/310 (97%)

Query: 1   MTYAGNRANIEAILGDRVELVVGDIADAELVDKLAAKADAIVHYAAESHNDNSLNDPSPF 60
            +TYAGN AN+E ILGDRVELVVGDIAD+ELVDKLAAKADAIVHYAAESHNDNSL DPSPF
Sbjct: 39  LTYAGNHANLEEILGDRVELVVGDIADSELVDKLAAKADAIVHYAAESHNDNSLKDPSPF 98

Query: 61  IHTNFIGTYTLLEAARKYDIRFHHVSTDEVYGDLPLREDLPGNGEGPGEKFTAETKYNPS 120
            I+TNF+GTYTLLEAARKYDIRFHHVSTDEVYGDLPLREDLPG+GEGPGEKFTAETKYNPS
Sbjct: 99  IYTNFVGTYTLLEAARKYDIRFHHVSTDEVYGDLPLREDLPGHGEGPGEKFTAETKYNPS 158

Query: 121 SPYSSTKAASDLIVKAWVRSFGVKATISNCSNNYGPYQHIEKFIPRQITNILAGIKPKLY 180
            SPYSSTKAASDLIVKAWVRSFGVKATISNCSNNYGPYQHIEKFIPRQITNIL+GIKPKLY
Sbjct: 159 SPYSSTKAASDLIVKAWVRSFGVKATISNCSNNYGPYQHIEKFIPRQITNILSGIKPKLY 218

Query: 181 GEGKNVRDWIHTNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKMGQPKDAY 240
```

```
                GEGKNVRDWIHTNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKM QPKDAY
Sbjct: 219 GEGKNVRDWIHTNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKMSQPKDAY 278

Query: 241 DHVTDRAGHDLRYAIDSTKLREELGWEPQFTNFSEGLEETINWYTENQDWWKAEKEAVEA 300
            DHVTDRAGHDLRYAIDSTKLREELGW+PQFTNF EGLE+TI WYTE++DWWKAEKEAVEA
Sbjct: 279 DHVTDRAGHDLRYAIDSTKLREELGWKPQFTNFEEGLEDTIKWYTEHEDWWKAEKEAVEA 338

Query: 301 NYAKTQEVIN 310
            NYAKTQ+++N
Sbjct: 339 NYAKTQKILN 348
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2673> which encodes the amino acid sequence <SEQ ID 2674>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1150(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 300/309 (97%), Positives = 303/309 (97%)

Query: 1   MTYAGNRANIEAILGDRVELVVGDIADAELVDKLAAKADAIVHYAAESHNDNSLNDPSPF 60
            +TYAGNRANIEAILGDRVELVVGDIADAELVDKLAAK DAIVHYAAESHNDNSL DPSPF
Sbjct: 37  LTYAGNRANIEAILGDRVELVVGDIADAELVDKLAAKTDAIVHYAAESHNDNSLEDPSPF 96

Query: 61  IHTNFIGTYTLLEAARKYDIRFHHVSTDEVYGDLPLREDLPGNGEGPGEKFTAETKYNPS 120
            IHTNFIGTYTLLEAARKYDIRFHHVSTDEVYGDLPLREDLPG GEGPGEKFTAETKYNPS
Sbjct: 97  IHTNFIGTYTLLEAARKYDIRFHHVSTDEVYGDLPLREDLPGQGEGPGEKFTAETKYNPS 156

Query: 121 SPYSSTKAASDLIVKAWVRSFGVKATISNCSNNYGPYQHIEKFIPRQITNILAGIKPKLY 180
            SPYSSTKAASDLIVKAWVRSFGVKATISNCSNNYGPYQHIEKFIPRQITNILAGIKPKLY
Sbjct: 157 SPYSSTKAASDLIVKAWVRSFGVKATISNCSNNYGPYQHIEKFIPRQITNILAGIKPKLY 216

Query: 181 GEGKNVRDWIHTNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKMGQPKDAY 240
            GEGKNVRDWIHTNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKMGQPKDAY
Sbjct: 217 GEGKNVRDWIHTNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKMGQPKDAY 276

Query: 241 DHVTDRAGHDLRYAIDSTKLREELGWEPQFTNFSEGLEETINWYTENQDWWKAEKEAVEA 300
            DHVTDRAGHDLRYAIDSTKLREELGWEPQFTNFSEGLEETI WYTEN+ WWKAEK+AVEA
Sbjct: 277 DHVTDRAGHDLRYAIDSTKLREELGWEPQFTNFSEGLEETIKWYTENETWWKAEKDAVEA 336

Query: 301 NYAKTQEVI 309
             YAKTQEVI
Sbjct: 337 KYAKTQEVI 345
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 881

A DNA sequence (GBSx0935) was identified in *S.agalactiae* <SEQ ID 2675> which encodes the amino acid sequence <SEQ ID 2676>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 882

A DNA sequence (GBSx0936) was identified in *S.agalactiae* <SEQ ID 2677> which encodes the amino acid sequence <SEQ ID 2678>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-15.55   Transmembrane   13 -  29 (   3 -  40)

----- Final Results -----
               bacterial membrane --- Certainty=0.7220(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 883

A DNA sequence (GBSx0937) was identified in *S.agalactiae* <SEQ ID 2679> which encodes the amino acid sequence <SEQ ID 2680>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2882(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 884

A DNA sequence (GBSx0938) was identified in *S.agalactiae* <SEQ ID 2681> which encodes the amino acid sequence <SEQ ID 2682>. This protein is predicted to be hyaluronate lyase. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
```

```
                    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2683> which encodes the amino acid sequence <SEQ ID 2684>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9099> which encodes the amino acid sequence <SEQ ID 9100>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 23
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                 bacterial outside --- Certainty= 0.300(Affirmative) < succ>
                bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 359/771 (46%), Positives = 492/771 (63%), Gaps = 50/771 (6%)

Query: 307  PNAT--GSTTVKISDKSGKIIKEVPLSVTASTEDNFTKLLDKWNDVTIGNHVYDTNDSNM 364
            PN T   + T+  +D   K+++          +D +T+LLD+WN +  GN  YD  + +M
Sbjct: 65   PNNTYFQTQTLTTTDSEKKVVQP-------QQKDYYTELLDQWNSIIAGNDAYDKTNPDM 117

Query: 365  QKLNQKLDETNAKNIEAIKL-----DSNRTFLWKDLDNLNNSAQLTATYRRLEDLAKQIT 419
               + K  E +A+NI  IK        NRT+LW+   + + SA +T TYR +E +AKQIT
Sbjct: 118  VTFHNKA-EKDAQNI--IKSYQGPDHENRTYLWEHAKDYSASANITKTYRNIEKIAKQIT 174

Query: 420  NPHSTIYKNEKAIRTVKESLAWLHQNFYNVNKDI------EGSANWWDFEIGVPRSITGT 473
            NP S  Y++ KAI  VK+ +A+++++ YN++++       E   NWW +EIG PR+I  T
Sbjct: 175  NPESCYYQDSKAIAIVKDGMAFMYEHAYNLDRENHQTTGKENKENWWVYEIGTPRAINNT 234

Query: 474  LALMYNYFTDAEIKTYTDPIEHFVPDAGFFRKTLVN--PFKALGGNLVDMGRVKIIEGLL 531
            L+LMY YFT  EI  YT PIE FVPD   FR    N  PF+A  GNL+DMGRVK+I G+L
Sbjct: 235  LSLMYPYFTQEEILKYTAPIEKFVPDPTRFRVRAANFSPFEANSGNLIDMGRVKLISGIL 294

Query: 532  RKDNTIIEKTSHSLKNLFTTATKAEGFYADGSYIDHT-----------NVAYTGAYGNVL 580
            RKD+  I  T  +++ +FT   +  GFY DGS IDH             +AYTGAYGNVL
Sbjct: 295  RKDDLEISDTIKAIEKVFTLVDEGNGFYQDGSLIDHVVTNAQSPLYKKGIAYTGAYGNVL 354

Query: 581  IDGLTQLLPIIQETDYKISNQELDMVYKWINQSFLPLIVKGELMDMSRGRSISREAASSH 640
            IDGL+QL+PIIQ+T   I   ++  +Y WIN SF P+IV+GE+MDM+RGRSISR  A SH
Sbjct: 355  IDGLSQLIPIIQKTKSPIKADKMATIYHWINHSFFPIIVRGEMMDMTRGRSISRFNAQSH 414

Query: 641  AAAVEVLRGFLRLANMSNEERNLDLKSTIKTIITS-NKFYNVFNNLKSYSDIANMNKLLN 699
             A +E LR  LR+A+MS E   L LK+ IKT++T  N FYNV++NLK+Y DI  M +LL+
Sbjct: 415  VAGIEALRAILRIADMSEEPHRLALKTRIKTLVTQGNAFYNVYDNLKTYHDIKLMKELLS 474

Query: 700  DSTVATKPLKSNLSTFNSMDRLAYYNAEKDFGFALSLHSKRTLNYEGMNDENTRGWYTGD 759
            D++V  + L S +++FNSMD+LA YN + DF F LS+ S RT NYE MN+EN  GW+T D
Sbjct: 475  DTSVPVQKLDSYVASFNSMDKLALYNNKHDFAFGLSMFSNRTQNYEAMNNENLHGWFTSD 534

Query: 760  GMFYLYNSDQSHYSNHFWPTVNPYKMAGTTEKDAKREDTTKDFMSKHSKDAKEKTGQVTG 819
            GMFYLYN+D  HYS ++W TVNPY++ GTTE + K  + T +    + K   ++ G +TG
Sbjct: 535  GMFYLYNNDLGHYSENYWATVNPYRLPGTTETEQKPLEGTPE----NIKTNYQQVG-MTG 589

Query: 820  ASD--FVGSVKLNDHFALAAMDFTNWDRTLTAQKGWVILNDKIVFLGSNIKNTNGIGNVS 877
```

```
             SD  FV S KLN+  ALAAM FTNW+++LT  KGW IL +KI+F+GSNIKN +
Sbjct: 590  LSDDAFVASKKLNNTSALAAMTFTNWNKSLTLNKGWFILGNKIIFVGSNIKNQSS-HKAY 648

Query: 878  TTIDQRKDDSKTPYTTYVNGKTVDLKQASSQQFTDTKSVFLESKEPGRNIGYIFFKNSTI 937
            TTI+QRK++ K PY +YVN + VDL       FT+TKS+FLES +P +NIGY FFK +T+
Sbjct: 649  TTIEQRKENQKYPYCSYVNNQPVDLNN-QLVDFTNTKSIFLESDDPAQNIGYYFFKPTTL 707

Query: 938  DIERKEQTGTWNSINRTSKNTSI---VSNPFITISQKHDNKGDSYDYMMVPNIDRTSFDK 994
             I +  QTG W +I    K+      VSN FITI Q H   GD Y YMM+PN+ R  F+
Sbjct: 708  SISKALQTGKWQNIKADDKSPEAIKEVSNTFITIMQNHTQDGDRYAYMMLPNMTRQEFET 767

Query: 995  LANSKEVELLENSSKQQVIYDKNSQTWAVIKHDNQESLINNQFKMNKAGLY 1045
              +  +++LLEN+ K   +YD +SQ   VI +  + ++ +N   ++  G Y
Sbjct: 768  YISKLDIDLLENNDKLAAVYDHDSQQMHVIHYGKKATMFSNH-NLSHQGFY 817
```

SEQ ID 2682 (GBS89) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 6 (lane 3; MW 118kDa).

The His-fusion protein was purified as shown in Figure 190, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 885

A DNA sequence (GBSx0939) was identified in *S.agalactiae* <SEQ ID 2685> which encodes the amino acid sequence <SEQ ID 2686>. This protein is predicted to be mutator mutt protein. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3781(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA11250 GB:D78182 MutX [Streptococcus mutans]
 Identities = 132/160 (82%), Positives = 146/160 (90%), Gaps = 1/160 (0%)

Query: 1   MTKLATICYIDNGKELLLLHRNKKENDVHEGKWISVGGKLEAGETPDECAKREILEETHL 60
           M KLATICYIDNG+ELLL+HRNKK NDVHEGKWISVGGKLE GE+PDECA+REI EETHL
Sbjct: 1   MIKLATICYIDNGRELLLMHRNKKPNDVHEGKWISVGGKLEKGESPDECARREIFEETHL 60

Query: 61  TVKKMDFKGVITFPEFTPGHDWYTYVFKVTDYEGELISDDESREGTLEWVPYDQVLSKPT 120
            VK+MDFKG+ITFP+FTPGHDWYTYVFKV D+EG LISD +SREGTLEWVPY+QVL+KPT
Sbjct: 61  IVKQMDFKGIITFPDFTPGHDWYTYVFKVRDFEGRLISDKDSREGTLEWVPYNQVLTKPT 120

Query: 121 WQGDYEIFKWILEDVPFFSAKFVYDEHQNLIEKTVNFYEK 160
           W+GDYEIFKWILED PFFSAKFVY E Q L++K V FYEK
Sbjct: 121 WEGDYEIFKWILEDAPFFSAKFVYQE-QKLVDKHVIFYEK 159
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2687> which encodes the amino acid sequence <SEQ ID 2688>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3399(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 131/158 (82%), Positives = 146/158 (91%)

Query: 1   MTKLATICYIDNGKELLLLHRNKKENDVHEGKWISVGGKLEAGETPDECAKREILEETHL 60
           MT+LATICYIDNG  LLLLHRNKKENDVH+GKWISVGGKLEAGETPDECA+REILEETHL
Sbjct: 1   MTQLATICYIDNGDSLLLLHRNKKENDVHKGKWISVGGKLEAGETPDECARREILEETHL 60

Query: 61  TVKKMDFKGVITFPEFTPGHDWYTYVFKVTDYEGELISDDESREGTLEWVPYDQVLSKPT 120
           TV +M FKG+ITFPEFTPGHDWYTYVFKVT +EG+LISD+ESREGTLEWVPYDQVL KPT
Sbjct: 61  TVTEMAFKGIITFPEFTPGHDWYTYVFKVTGFEGDLISDEESREGTLEWVPYDQVLEKPT 120

Query: 121 WQGDYEIFKWILEDVPFFSAKFVYDEHQNLIEKTVNFY 158
           W+GDY+IFKWILED  FFSAKF YD++  L++K+V FY
Sbjct: 121 WEGDYDIFKWILEDRSFFSAKFTYDQNNQLMDKSVTFY 158
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 886

A DNA sequence (GBSx0940) was identified in *S.agalactiae* <SEQ ID 2689> which encodes the amino acid sequence <SEQ ID 2690>. This protein is predicted to be MutT/nudix family protein. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1901(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF11817 GB:AE002059 MutT/nudix family protein [Deinococcus radiodurans]
 Identities = 40/135 (29%), Positives = 62/135 (45%), Gaps = 3/135 (2%)

Query: 22  FGVRVSALIIENQKLLLIYAPHLDKYY-LPGGALQVGEDSNKAVAREVLEEIGLHSQVGD 80
           F  R + + +++ +LL  +      ++ LPGGA+Q GE S  A  RE  EE GL + V
Sbjct: 33  FQTRATLICVQDNRLLTCWDERFPDFFALPGGAVQTGESSAAAAQREWHEETGLRADVTR 92

Query: 81  LAYIIENQFNIKRHHYHSVEFLYFVNLLGQAPESIKEGTHKRHFVWLPIKELTKIDCNPN 140
            A  +E  F+ +    H   F + V L G+ P ++ +  H   F WL +  L      P
Sbjct: 93  CA-TLERFFHWEGRERHEFGFFFRVELTGELPATVLDNPHV-FFRWLAVDALDDHTLYPR 150

Query: 141 FLAQDLIEWPGHVVH 155
            + Q L    G + H
Sbjct: 151 CVPQLLRLPAGEIGH 165
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2691> which encodes the amino acid sequence <SEQ ID 2692>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3832(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 33/80 (41%), Positives = 50/80 (62%), Gaps = 1/80 (1%)

Query: 29  LIIENQKLLLIYAPHLDKYYLPGGALQVGEDSNKAVAREVLEEIGLHSQVGDLAYIIENQ 88
            LI+ N K  L      D+YY  GG   VGE +++ V RE LEE+G+ ++V  LA+++EN
Sbjct: 1   LIVRNGKNFLTRDAD-DQYYTIGGTSLVGEKTHETVLRETLEEVGIRAKVNQLAFMVENH 59

Query: 89  FNIKRHHYHSVEFLYFVNLL 108
            F+I    +H++EF Y V+ L
Sbjct: 60  FDIDDVFWHNIEFHYLVSPL 79
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 887

A DNA sequence (GBSx0941) was identified in *S.agalactiae* <SEQ ID 2693> which encodes the amino acid sequence <SEQ ID 2694>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.95    Transmembrane   24 -  40 (  17 -  48)
   INTEGRAL    Likelihood =-11.09    Transmembrane   88 - 104 (  82 - 112)
   INTEGRAL    Likelihood = -9.39    Transmembrane  294 - 310 ( 276 - 315)
   INTEGRAL    Likelihood = -8.07    Transmembrane  242 - 258 ( 236 - 262)
   INTEGRAL    Likelihood = -7.86    Transmembrane   50 -  66 (  43 -  74)
   INTEGRAL    Likelihood = -3.13    Transmembrane  337 - 353 ( 332 - 355)
   INTEGRAL    Likelihood = -2.23    Transmembrane  185 - 201 ( 182 - 202)
   INTEGRAL    Likelihood = -1.38    Transmembrane  269 - 285 ( 267 - 285)

----- Final Results -----
            bacterial membrane --- Certainty=0.6180(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2695> which encodes the amino acid sequence <SEQ ID 2696>. Analysis of this protein sequence reveals the following:

```
     Possible site: 26
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.71    Transmembrane   88 - 104 (  85 - 112)
   INTEGRAL    Likelihood = -9.29    Transmembrane   24 -  40 (  21 -  72)
   INTEGRAL    Likelihood = -8.92    Transmembrane   47 -  63 (  41 -  72)
   INTEGRAL    Likelihood = -7.59    Transmembrane  243 - 259 ( 237 - 266)
   INTEGRAL    Likelihood = -6.10    Transmembrane  181 - 197 ( 178 - 203)
   INTEGRAL    Likelihood = -5.47    Transmembrane  278 - 294 ( 273 - 310)
   INTEGRAL    Likelihood = -3.88    Transmembrane  338 - 354 ( 331 - 368)
   INTEGRAL    Likelihood = -1.59    Transmembrane  297 - 313 ( 297 - 314)

----- Final Results -----
            bacterial membrane --- Certainty=0.4885(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD00285 GB:U78604 putative membrane protein [Streptococcus mutans]
 Identities = 244/382 (63%), Positives = 310/382 (80%), Gaps = 3/382 (0%)

Query: 12  SLFYKWFLNNQATMALVITLLAFLTIFVFTKISFLFMPVISFFAVIMLPLVISTILYYLT 71
            S F+KWFL+N+    L++ LL FL I VFTKIS +F P++SF AVIMLPLVIS +LYYL
Sbjct: 17  SWFFKWFLDNKTVTVLLVLLLVFLDILVFTKISSIFKPLLSFLAVIMLPLVISALLYYLL 76

Query: 72  KPLVDLINHLGPNRTTSIFIVFGLITLLFVWAISGFVPMVQTQLTSFIEDLPKYVGKVNE 131
```

```
           KP+VD I   G +R  +I IVF +I  L VW I+ F PM+  QLTSFI+ LP YV  V+
Sbjct: 77  KPIVDFIEIRGTSRVMAITIVFVIIAGLLVWGIANFFPMLNEQLTSFIKYLPSYVRSVDA 136

Query: 132 EANKLLENEWLVSYKPQLQDMLTHTSQKALDYAQSFSKNAIDWAGNFAGAIARITVAIII 191
           + +KLL N+ L S++PQ+++ +T+ SQKA+DYA+ FSK A+ WAGNFA  IAR+TVAIII
Sbjct: 137 QVSKLLRNDLLASFRPQIENAVTNFSQKAVDYAEPFSKGAVTWAGNFASLIARVTVAIII 196

Query: 192 SPFILFYFLRDSSHMKNGLVNVLPLKLRVPMVRVLGDINKQLSGYVQGQVTVAIVVGFMF 251
           SPFI+FY LRDSS MK   V+ LP K+R P+ R+LGD+N+QL+GYVQ   TVAI+VGFMF
Sbjct: 197 SPFIVFYLLRDSSKMKEAFVSYLPTKMRQPIHRILGDVNRQLAGYVQRSSTVAIIVGFMF 256

Query: 252 SIMFSLVGLKYAITFGIIAGFLNMIPYLGSFLAMIPVVIMAMVQGPFMLVKVLVIFMIEQ 311
           SIMF+++GL+YA+TFGIIAGFLNMIPYLGSFLA IPV I+A+V+GP  +VKV ++F++EQ
Sbjct: 257 SIMFTIIGLRYAVTFGIIAGFLNMIPYLGSFLATIPVFILALVEGPVKVVKVALVFIVEQ 316

Query: 312 TIEGRFVAPLVLGNKLSIHPITIMFLLLTAGSMFGVWGVFLVIPIYASVKVVIKELFDWY 371
           TIEGRFV+PLVLG+KLSIHPITIMF+LLTAGSMFGVWGVFL IP+YAS+KVV+KE+F+WY
Sbjct: 317 TIEGRFVSPLVLGSKLSIHPITIMFILLTAGSMFGVWGVFLGIPVYASIKVVVKEIFEWY 376

Query: 372 KKVSGLYDEEVLVIEEVKDHVK 393
           K +SGLY++E    E++K  VK
Sbjct: 377 KPISGLYEKEE---EDIKKDVK 395
```

An alignment of the GAS and GBS proteins is shown below.

```
  Identities = 243/389 (62%), Positives = 306/389 (78%), Gaps = 2/389 (0%)

Query: 6   EKEFKNSLFFKWILNNQAVIALMITFLVFLTIFIFTKISFMFKPVFDFLAVLILPLVISG 65
           EK   +SLF+KW LNNQA +AL+IT L FLTIF+FTKISF+F PV  F AV++LPLVIS
Sbjct: 6   EKSRTDSLFYKWFLNNQATMALVITLLAFLTIFVFTKISFLFMPVISFFAVIMLPLVIST 65

Query: 66  LLYYLLKPMVTFLEKRGIKRVTAILSVFTIIILLLIWAMSSFIPMMSNQLRHFMEDLPSY 125
           +LYYL KP+V  +   G  R T+I  VF +I LL +WA+S F+PM+  QL  F+EDLP Y
Sbjct: 66  ILYYLTKPLVDLINHLGPNRTTSIFIVFGLITLLFVWAISGFVPMVQTQLTSFIEDLPKY 125

Query: 126 VNKVQMETSSFIDHNPWLKSYKGEISSMLSNISSQAVSYAEKFSKNILDWAGNLASTVAR 185
           V KV  E +  ++ N WL SYK ++  ML++ S +A+ YA+ FSKN +DWAGN A  +AR
Sbjct: 126 VGKVNEEANKLLE-NEWLVSYKPQLQDMLTHTSQKALDYAQSFSKNAIDWAGNFAGAIAR 184

Query: 186 VTVATIMAPFILFYLLRDSRNMKNGFLMVLPTKLRQPTDRILREMNSQMSGYVQGQIIVA 245
           +TVA I++PFILFY LRDS +MKNG + VLP KLR P  R+L ++N Q+SGYVQGQ+ VA
Sbjct: 185 ITVAIIISPFILFYFLRDSSHMKNGLVNVLPLKLRVPMVRVLGDINKQLSGYVQGQVTVA 244

Query: 246 ITVGVIFSIMYSIIGLRYGVTLGIIAGVLNMVPYLGSFVAQIPVFILALVAGPVMVVKVA 305
           I VG +FSIM+S++GL+Y +T GIIAG LNM+PYLGSF+A IPV I+A+V GP M+VKV
Sbjct: 245 IVVGFMFSIMFSLVGLKYAITFGIIAGFLNMIPYLGSFLAMIPVVIMAMVQGPFMLVKVL 304

Query: 306 IVFVIEQTLEGRFVSPLVLGNKLSIHPITIMFILLTSGAMFGVWGVFLSIPIYASIKVVV 365
           ++F+IEQT+EGRFV+PLVLGNKLSIHPITIMF+LLT+G+MFGVWGVFL IPIYAS+KVV+
Sbjct: 305 VIFMIEQTIEGRFVAPLVLGNKLSIHPITIMFLLLTAGSMFGVWGVFLVIPIYASVKVVI 364

Query: 366 KELFDWYKAVSGLYTVDV-VTEERSEEVK 393
           KELFDWYK VSGLY  +V V EE  + VK
Sbjct: 365 KELFDWYKKVSGLYDEEVLVIEEVKDHVK 393
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 888

A DNA sequence (GBSx0942) was identified in *S.agalactiae* <SEQ ID 2697> which encodes the amino acid sequence <SEQ ID 2698>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2715(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9891> which encodes amino acid sequence <SEQ ID 9892> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAA25160 GB:L16975 ORF1 [Lactococcus lactis]
 Identities = 132/345 (38%), Positives = 203/345 (58%), Gaps = 3/345 (0%)

Query: 79  INLAQIVAEDGDIEQAFLYLDYISEDSQEYVSALLVMADLYDMEGLTDVAREKLLLASKL 138
            +NLA+I  ++G++++A  YL  I  + + Y++AL+ +ADLY  E   + A  KL  A +L
Sbjct: 1   VNLAEIAEDNGNLDEALNYLYQIPVNDENYIAALIKIADLYQFEVDFETAISKLEEAREL 60

Query: 139 SDDPLVTFGLAEMNLSLEHYQEAIEGYASLDNREILETTGVSTYQRIGKSYAIMGKFDAA 198
            SD PL+TF LAE       Y  AI  YA L  R+IL  T +S YQRIG SYA +G F+ A
Sbjct: 61  SDSPLITFALAESYFEQGDYSAAITEYAKLSERKILHETKISIYQRIGDSYAQLGNFENA 120

Query: 199 IEFLEKAVDIEYDDLTVFELATILYDQEEYQKANLYFKQLDTINPDFAGYEYIYGLSLRE 258
            I FLEK+++ +     T++++A +  +     +A   FK+L+ ++ +F  YE  Y  +L
Sbjct: 121 ISFLEKSLEFDEKPETLYKIALLYGETHNETRAIANFKRLEKMDVEFLNYELAYAQTLEA 180

Query: 259 EHKSEEALRLVQQGIRKNSFDGQLLLLASQLSYELHDVHSSESYLKQAEKVSENQDEIVM 318
             + + AL + ++G++KN     LL  AS++ ++L D  ++E YL  A + E  DE V
Sbjct: 181 NQEFKAALEMAKKGMKKNPNAVPLLHFASKICFKLKDKAAAERYLVDALNLPELHDETVF 240

Query: 319 RLSNLYLEEERFEEVLELDN-DNLENILAKWNIAKAHKALEMDDSVD--YYQSLYNDLKD 375
             L+NLY  EE FE V+ L+     E++LAKW  A AHKALE D      Y + +  +L +
Sbjct: 241 LLANLYFNEEDFEAVINLEELLEDEHLLAKWLFAGAHKALENDSEAAALYEELIQTNLSE 300

Query: 376 NPEFLQDYAYILREFGYLDKAQEVGKAYLKLVPDDIEMSEWVNNI 420
            NPEFL+DY   L+E G + K + + + YL+LVPDD  M   + ++
Sbjct: 301 NPEFLEDYIDFLKEIGQISKTEPIIEQYLELVPDDENMRNLLTDL 345
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2699> which encodes the amino acid sequence <SEQ ID 2700>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2991(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 267/409 (65%), Positives = 336/409 (81%), Gaps = 1/409 (0%)

Query: 13  MLNSEKMIVSIQNQDLEHANKYFEKALKNDPEEVLLELGAYLESIGFLPQAKRLYDQIRP 72
            MLNSEKMI S+  QDL HA KYF+KALK D  + L+ LG YLESIGFLP AKR+Y Q+
Sbjct: 7   MLNSEKMIASLDQQDLAHAEKYFQKALKEDDADSLIALGEYLESIGFLPHAKRIYLQLAD 66

Query: 73  NYPEVAINLAQIVAEDGDIEQAFLYLDYISEDSQEYVSALLVMADLYDMEGLTDVAREKL 132
            +YPE+ INLAQI AED  IE+AFLYLD +S+DS  Y+SALLVMADLYDMEGLT+VAREKL
Sbjct: 67  DYPELNINLAQIAAEDDAIEEAFLYLDKVSKDSPNYLSALLVMADLYDMEGLTEVAREKL 126

Query: 133 LLASKLSDDPLVTFGLAEMNLSLEHYQEAIEGYASLDNREILETTGVSTYQRIGKSYAIM 192
            L A  +S +PLV FGLAE+++SL+H++EAI+ YA LDNR+ILE TG+STYQRIG++YA +
Sbjct: 127 LQAVGISPEPLVIFGLAEIDMSLQHFKEAIDYYAQLDNRQILELTGISTYQRIGRAYASL 186

Query: 193 GKFDAAIEFLEKAVDIEYDDLTVFELATILYDQEEYQKANLYFKQLDTINPDFAGYEYIY 252
            GKF+AAIEFLEKAV IEY+D TVFELAT++YDQE YQKANLYFKQL+TINPD+ GYEY Y
```

```
Sbjct: 187 GKFEAAIEFLEKAVAIEYEDETVFELATLMYDQENYQKANLYFKQLETINPDYPGYEYGY 246

Query: 253 GLSLREEHKSEEALRLVQQGIRKNSFDGQLLLLASQLSYELHDVHSSESYLKQAEKVSEN 312
            LSL EEHK+ EALRLVQQG+RKN+FD QLLLLASQLSYELHD  ++E+YL QA++V+ +
Sbjct: 247 ALSLHEEHKTSEALRLVQQGLRKNAFDSQLLLLASQLSYELHDRQNAENYLLQAKEVAVD 306

Query: 313 QDEIVMRLSNLYLEEERFEEVLELDNDNLENILAKWNIAKAHKALEMDD-SVDYYQSLYN 371
            +EI+MRL  LY + ERFEEV+ L+ + ++N+L KW IAKA+ ALE ++ ++  Y  +
Sbjct: 307 DEEILMRLVTLYFDAERFEEVIALNRETIDNVLTKWTIAKAYHALEQEEVALALYNEISA 366

Query: 372 DLKDNPEFLQDYAYILREFGYLDKAQEVGKAYLKLVPDDIEMSEWVNNI 420
           DL +NPEFLQDYAY+LREFG   KA ++  AYL+ VPDD+ M +++++I
Sbjct: 367 DLAENPEFLQDYAYLLREFGQFHKAIQMATAYLRQVPDDVNMQDFLDHI 415
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 889

A DNA sequence (GBSx0943) was identified in *S.agalactiae* <SEQ ID 2701> which encodes the amino acid sequence <SEQ ID 2702>. This protein is predicted to be alpha-acetolactate synthase (ilvK). Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2105(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA01700 GB:A23961 alpha-acetolactate synthase [Lactococcus
           lactis]
 Identities = 396/559 (70%), Positives = 466/559 (82%), Gaps = 8/559 (1%)

Query: 4   SHNQYGADLIVDSLINHDVKYVFGIPGAKIDRVFDTLE-DKGPELIVARHEQNATFMAQA 62
           S  Q+GA+L+VDSLINH VKYVFGIPGAKIDRVFD LE ++GP+++V RHEQ A FMAQA
Sbjct: 2   SEKQFGANLVVDSLINHKVKYVFGIPGAKIDRVFDLLENEEGPQMVVTRHEQGAAFMAQA 61

Query: 63  VGRITGEPGVVIATSGPGISNLATGLVTATDEGDAVLAIGGQVKRGDLLKRAHQSMNNVA 122
           VGR+TGEPGVV+ TSGPG+SNLAT L+TAT EGDA+LAIGGQVKR D LKRAHQSM+N
Sbjct: 62  VGRLTGEPGVVVVTSGPGVSNLATPLLTATSEGDAILAIGGQVKRSDRLKRAHQSMDNAG 121

Query: 123 MLEPITKYSAEVHDPNTLSETVANAYRLAKSGKPGASFISIPQDVTDSPVSVKAIKPLSA 182
           M++  TKYSAEV DPNTLSE++ANAYR+AKSG PGA+F+SIPQDVTD+ VS+KAI+PLS
Sbjct: 122 MMQSATKYSAEVLDPNTLSESIANAYRIAKSGHPGATFLSIPQDVTDAEVSIKAIQPLSD 181

Query: 183 PKLGSASVLDINYLAQAINNAVLPVLLLGNGASSEGVTAAVRRLLDAVKLPVVETFQGAG 242
           PK+G+AS+ DINYLAQAI NAVLPV+L+G GAS   V +++R LL  V +PVVETFQGAG
Sbjct: 182 PKMGNASIDDINYLAQAIKNAVLPVILVGAGASDAKVASSLRNLLTHVNIPVVETFQGAG 241

Query: 243 IVSRELEDETFFGRVGLFRNQPGDMLLKRADLVIAIGYDPIEYEARNWNAEISARIIVID 302
           ++S +LE  TF+GR+GLFRNQPGDMLLKR+DLVIA+GYDPIEYEARNWNAEI +RIIVID
Sbjct: 242 VISHDLE-HTFYGRIGLFRNQPGDMLLKRSDLVIAVGYDPIEYEARNWNAEIDSRIIVID 300

Query: 303 VEQAEIDTYFQPERELIGDMAHTLDLLLPAIKGYELPEGSKEYLKGLRNNIENVSDVKFD 362
              AEIDTY+QPERELIGD+A TLD LLPA++GY++P+G+K+YL GL    E     +FD
Sbjct: 301 NAIAEIDTYYQPERELIGDIAATLDNLLPAVRGYKIPKGTKDYLDGLH---EVAEQHEFD 357

Query: 363 RDSA-HGLVHPLDLIDVLQENTTDDMTVTVDVGSHYIWMARYFKSYEARHLLFSNGMQTL 421
            ++   G +HPLDL+    QE   DD TVTVDVGS YIWMAR+FKSYE RHLLFSNGMQTL
Sbjct: 358 TENTEEGRMHPLDLVSTFQEIVKDDETVTVDVGSLYIWMARHFKSYEPRHLLFSNGMQTL 417
```

```
Query: 422 GVALPWAISAALLRPNTKVISVSGDGGFLFSAQELETAVRLHLPIVHIIWNDGKYNMVEF 481
           GVALPWAI+AALLRP  KV S SGDGGFLF+ QELETAVRL+LPIV IIWNDG Y+MV+F
Sbjct: 418 GVALPWAITAALLRPGKKVYSHSGDGGFLFTGQELETAVRLNLPIVQIIWNDGHYDMVKF 477

Query: 482 QEEMKYGRSSGVDFGPVDFVKYAESFGAKGYRVDSKDSFEETLKQALIDAENGPVLIDVP 541
           QEEMKYGRS+ VDFG VD+VKYAE+  AKGYR  SK+   E LK   I    GPV+IDVP
Sbjct: 478 QEEMKYGRSAAVDFGYVDYVKYAEAMRAKGYRAHSKEELAEILKS--IPDTTGPVVIDVP 535

Query: 542 IDYKDNVTLGETILPDEFY 560
           +DY DN+ L E +LP+EFY
Sbjct: 536 LDYSDNIKLAEKLLPEEFY 554
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 890

A DNA sequence (GBSx0944) was identified in *S.agalactiae* <SEQ ID 2703> which encodes the amino acid sequence <SEQ ID 2704>. This protein is predicted to be alpha-acetolactate decarboxylase (aldC). Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3096(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9889> which encodes amino acid sequence <SEQ ID 9890> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA57941 GB:X82620 alpha-acetolactate decarboxylase [Lactococcus
           lactis]
 Identities = 139/239 (58%), Positives = 187/239 (78%), Gaps = 3/239 (1%)

Query: 16  MSETVKLFQYSTLSSLMAGLYKGSLTIGELLTHGDLGIGTVHMIDGELIVLDGKAYQAIG 75
           MSE  +LFQY+TL +LMAGLY+G++TIGELL HGDLGIGT+  IDGELIVLDGKAYQA
Sbjct: 1   MSEITQLFQYNTLGALMAGLYEGTMTIGELLKHGDLGIGTLDSIDGELIVLDGKAYQA-- 58

Query: 76  TDGKAEIIQLSDDVTVPYAAVLPHHIQKQFDINAEIDNKDLEEMILKNFEGQNLFKSLKI 135
             G   I++L+DD+ VPYAAV+PH  +  F     + +K+LE+ I   F+GQNLF+S+KI
Sbjct: 59  -KGDKTIVELTDDIKVPYAAVVPHQAEVVFKQKFTVSDKELEDRIESYFDGQNLFRSIKI 117

Query: 136 KGTFSRMHVRMIPKSPQHKRFADIASNQPEFTRENVSGTLVGIWTPELFHGVGVKGFHVH 195
            G F +MHVRMIP++    +F +++ NQPE+T EN+ GT+VGIWTPE+FHGV V G+H+H
Sbjct: 118 TGKFPKMHVRMIPRAKSGTKFVEVSQNQPEYTEENIKGTIVGIWTPEMFHGVSVAGYHLH 177

Query: 196 FISDDLTFGGHVMDYSLTQGKVEIGKVDQLDQCFPTQDQEFLKANFDLQKLREDIDLSE 254
           FIS+D TFGGHV+D+ +  G VEIG +DQL+Q FP QD++FL A+ D++ L++DID++E
Sbjct: 178 FISEDFTFGGHVLDFIIDNGTVEIGAIDQLNQSFPVQDRKFLFADLDIEALKKDIDVAE 236
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 891

A DNA sequence (GBSx0945) was identified in *S.agalactiae* <SEQ ID 2705> which encodes the amino acid sequence <SEQ ID 2706>. This protein is predicted to be fibronectin-binding protein-like protein A. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5042(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA46282 GB:X65164 fibronectin-binding protein-like protein A
           [Streptococcus gordonii]
 Identities = 392/550 (71%), Positives = 462/550 (83%)

Query: 1   MSFDGFFLHHLTNELQEQIEKGRIQKVNQPFDHELVLTIRNNRRNYKLLLSAHPVFGRIQ 60
           MSFDGFFLHH+T EL+ ++  GRIQK+NQPF+ ELVL IR+NR++ KLLLSAH VFGR+Q
Sbjct: 1   MSFDGFFLHHMTEELRHELVGGRIQKINQPFEQELVLQIRSNRKSLKLLLSAHSVFGRVQ 60

Query: 61  TTEANFQNPQNPNTFTMIMRKYLQGAVIETIQQIENDRILEIVVSNKNEIGDHIKATLVV 120
            T+  F+NP  PNTF M+MRKYLQGAVIE IQQ+ENDRILEI VSNKNEIGD +  TLV+
Sbjct: 61  LTDTTFENPAVPNTFIMVMRKYLQGAVIEAIQQVENDRILEISVSNKNEIGDSVAVTLVI 120

Query: 121 EIMGKHSNIILIDKNEHKIIESIKHVGFSQNSYRTILPGSTYIAPPKTKAINPFDISDQT 180
           EIMGKHSNIIL+DK   KIIE+IKHVGFSQNSYRTILPGSTY+APP+T ++NPF + D+
Sbjct: 121 EIMGKHSNIILLDKASGKIIEAIKHVGFSQNSYRTILPGSTYVAPPQTGSLNPFTVGDEK 180

Query: 181 LFELLQTNDLSPKNLQQLLQGLGRDTALELSHCLKDNKLNDFRQFFSREYYPSLTEKSFS 240
           LFE+LQT ++ PK L Q+ QGLGRDTA ELS  L  ++L  FR FF+     PSLTEKSFS
Sbjct: 181 LFEILQTEEIEPKRLLQIFQGLGRDTATELSGRLTTDRLKTFRAFFASPTQPSLTEKSFS 240

Query: 241 AVQFSSSHETFQSLGQLLDYYYQEKAEKDRIAQQASDLIHRVQSELEKNIKKLAKQQDEL 300
           A+ FS S     +L +LLD +Y++KAE+ R+ QQAS+LI RV++ELEKN KKL KQ+DEL
Sbjct: 241 ALVFSDSKTQMSTLSELLDTFYKDKAERYRVNQQASELIRRVENELEKNRKKLGKQEDEL 300

Query: 301 LATENAEEFRQKGELLTTYLSMVPNNQDVVVLDNYYTNQTIEISLDRALTPNQNAQRYFK 360
           LATE AEEFRQKGELLTT+L  VPN+QD V LDNYYT + I I+LD+ALTPNQNAQRYFK
Sbjct: 301 LATEKAEEFRQKGELLTTFLHQVPNDQDQVELDNYYTGEKILITLDKALTPNQNAQRYFK 360

Query: 361 KYQKLKEAVKHLKGIISDTENTITYLESVETSLNHASMEDINDIREELVETGFIKRRAHD 420
           +YQKLKEAVKHL  +I +T  TI YLESVET+L  AS+ +I +IREEL++TGFI+RR  +
Sbjct: 361 RYQKLKEAVKHLTSLIEETRTTILYLESVETALAQASLTEIAEIREELIQTGFIRRRQRE 420

Query: 421 KQHKRKKPEQYLASDGKTIIMVGRNNLQNDELTFKMARKGELWFHAKDIPGSHVLIRDNL 480
           K  KRKKPE+YLASDG+TII+VGRNNLQNDELTFKMA+K ELWFHAKDIPGSHV+I  NL
Sbjct: 421 KIQKRKKPEKYLASDGQTIILVGRNNLQNDELTFKMAKKDELWFHAKDIPGSHVVITGNL 480

Query: 481 NPSDEVKTDAAELAAYYSKARLSNLVQVDMIEAKKLNKPSGTKPGFVTYTGQKTLRVTPT 540
            PSDEVKTDAAELAAY+SKARLSNLVQVDMIE KKLNKP+G KPGFVTYTGQKTLRVTP
Sbjct: 481 QPSDEVKTDAAELAAYFSKARLSNLVQVDMIEIKKLNKPTGGKPGFVTYTGQKTLRVTPD 540

Query: 541 QEKIDSLKLK 550
            +KI S+K++
Sbjct: 541 ADKIKSMKIQ 550
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2707> which encodes the amino acid sequence <SEQ ID 2708>. Analysis of this protein sequence reveals the following:

```
     Possible site: 38

>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
               bacterial cytoplasm --- Certainty=0.5434(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein differs significantly from L28919 in its mid-region:

```
Query: 223 QHFQGLGRDTAKELAELLTTD
             F  L  +T K + ELLTTD
Sbjct: 121 PAFSRLRGETPKRIGELLTTD
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 421/549 (76%), Positives = 487/549 (88%)

Query: 1   MSFDGFFLHHLTNELQEQIEKGRIQKVNQPFDHELVLTIRNNRRNYKLLLSAHPVFGRIQ 60
           MSFDGFFLHHLTNEL+E +  GRIQKVNQPF+ ELVLTIRN+R+NYKLLLSAHPVFGR+Q
Sbjct: 27  MSFDGFFLHHLTNELKENLLYGRIQKVNQPFERELVLTIRNHRKNYKLLLSAHPVFGRVQ 86

Query: 61  TTEANFQNPQNPNTFTMIMRKYLQGAVIETIQQIENDRILEIVVSNKNEIGDHIKATLVV 120
            T+A+FQNPQ PNTFTMIMRKYLQGAVIE ++QI+NDRI+EI VSNKNEIGD I+ATL++
Sbjct: 87  ITQADFQNPQVPNTFTMIMRKYLQGAVIEQLEQIDNDRIIEIKVSNKNEIGDAIQATLII 146

Query: 121 EIMGKHSNIILIDKNEHKIIESIKHVGFSQNSYRTILPGSTYIAPPKTKAINPFDISDQT 180
           EIMGKHSNIIL+D+ E+KIIESIKHVGFSQNSYRTILPGSTYI PPKT A+NPF I+D
Sbjct: 147 EIMGKHSNIILVDRAENKIIESIKHVGFSQNSYRTILPGSTYIEPPKTAAVNPFTITDVP 206

Query: 181 LFELLQTNDLSPKNLQQLLQGLGRDTALELSHCLKDNKLNDFRQFFSREYYPSLTEKSFS 240
           LFE+LQT +L+ K+LQQ  QGLGRDTA EL+  L  +KL  FR+FF+R    +LT  SF+
Sbjct: 207 LFEILQTQELTVKSLQQHFQGLGRDTAKELAELLTTDKLKRFREFFARPTQANLTTASFA 266

Query: 241 AVQFSSSHETFQSLGQLLDYYYQEKAEKDRIAQQASDLIHRVQSELEKNIKKLAKQQDEL 300
            V FS SH TF++L  +LD++YQ+KAE+DRI QQASDLIHRVQ+EL+KN  KL+KQ+ EL
Sbjct: 267 PVLFSDSHATFETLSDMLDHFYQDKAERDRINQQASDLIHRVQTELDKNRNKLSKQEAEL 326

Query: 301 LATENAEEFRQKGELLTTYLSMVPNNQDVVVLDNYYTNQTIEISLDRALTPNQNAQRYFK 360
           LATENAE FRQKGELLTTYLS+VPNNQD V+LDNYYT + IEI+LD+ALTPNQNAQRYFK
Sbjct: 327 LATENAELFRQKGELLTTYLSLVPNNQDSVILDNYYTGEKIEIALDKALTPNQNAQRYFK 386

Query: 361 KYQKLKEAVKHLKGIISDTENTITYLESVETSLNHASMEDINDIREELVETGFIKRRAHD 420
           KYQKLKEAVKHL G+I+DT+ +ITY ESV+ +L+ AS++DI DIREEL + GF+K R  D
Sbjct: 387 KYQKLKEAVKHLSGLIADTKQSITYFESVDYNLSQASIDDIEDIREELYQAGFLKSRQRD 446

Query: 421 KQHKRKKPEQYLASDGKTIIMVGRNNLQNDELTFKMARKGELWFHAKDIPGSHVLIRDNL 480
           K+HKRKKPEQYLASDG TI+MVGRNNLQN+ELTFKMA+KGELWFHAKDIPGSHV+I+DNL
Sbjct: 447 KRHKRKKPEQYLASDGTTILMVGRNNLQNEELTFKMAKKGELWFHAKDIPGSHVIIKDNL 506

Query: 481 NPSDEVKTDAAELAAYYSKARLSNLVQVDMIEAKKLNKPSGTKPGFVTYTGQKTLRVTPT 540
           +PSDEVKTDAAELAAYYSKARLSNLVQVDMIEAKKL+KPSG KPGFVTYTGQKTLRVTP
Sbjct: 507 DPSDEVKTDAAELAAYYSKARLSNLVQVDMIEAKKLHKPSGAKPGFVTYTGQKTLRVTPD 566

Query: 541 QEKIDSLKL 549
           Q KI S+KL
Sbjct: 567 QAKILSMKL 575
```

SEQ ID 2706 (GBS81) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 6 (lane 2; MW 64kDa) and in Figure 6 (lane 5; MW 64kDa). The GBS81-His fusion product was purified (Figure 190, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 319), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 892

A DNA sequence (GBSx0946) was identified in *S.agalactiae* <SEQ ID 2709> which encodes the amino acid sequence <SEQ ID 2710>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -9.08   Transmembrane    6 -  22 (   1 -  24)

----- Final Results -----
               bacterial membrane --- Certainty=0.4630(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF94260 GB:AE004191 conserved hypothetical protein [Vibrio cholerae]
 Identities = 111/295 (37%), Positives = 184/295 (61%), Gaps = 1/295 (0%)

Query: 36  QVVKIGILQYVTHDALDAIEKGVEDGLAQEGYK-GKKVKLTVLNAEADQSKIQAMSKQLV 94
           +  K+ + Q V H ALDA  +G+ DGL  +GY+ GK ++     A+ + +    +++Q V
Sbjct: 26  KTAKVAVSQIVEHPALDATRQGLLDGLKAKGYEEGKNLEFDYKTAQGNPAIAVQIARQFV 85

Query: 95  NHHNDILIGIATPSAQGLAASTKDTPIIMGAVSDPLGAKLVTNMKKPTTNVTGLSNVVPT 154
             + D+L+GIATP+AQ L ++TK  PI+  AV+DP+GAKLV  +++P  NVTGLS++ P
Sbjct: 86  GENPDVLVGIATPTAQALVSATKTIPIVFTAVTDPVGAKLVKQLEQPGKNVTGLSDLSPV 145

Query: 155 KQTVQLIKDITPNIKRIGILYASSEDNSVSQVTEFTKYAQKAGLEVLKYSVPSTNEIKTS 214
           +Q V+LIK+I PN+K IG++Y   E N+VS +      A K G+++++ +   + +++++
Sbjct: 146 EQHVELIKEILPNVKSIGVVYNPGEANAVSLMELLKLSAAKHGIKLVEATALKSADVQSA 205

Query: 215 MSVMTKKVDAVFVPQDNTIASAFRTVIVAANQANIPVYSSVDTMVEQGSIASVAQSQYGL 274
              + +K D ++   DNT+ASA   +IVAANQA  PV+ +  + VE+G+IAS+    Y +
Sbjct: 206 TQAIAEKSDVIYALIDNTVASAIEGMIVAANQAKTPVFGAATSYVERGAIASLGFDYYQI 265

Query: 275 GLETAKQAIKVLRGKPVKDVPVKVIDTGKPSLNLKAAKHLGIKIPKKIMKQAEIT 329
           G++TA     +L GK    + V+V       +N  AA+ LGI IP+ ++ +A  T
Sbjct: 266 GVQTADYVAAILEGKEPGSLDVQVAKGSDLVINKTAAEQLGITIPEAVLARATST 320
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2711> which encodes the amino acid sequence <SEQ ID 2712>. Analysis of this protein sequence reveals the following:

```
     Possible site: 23

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.25   Transmembrane    6 -  22 (   1 -  27)

----- Final Results -----
               bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF94260 GB:AE004191 conserved hypothetical protein [Vibrio cholerae]
 Identities = 103/304 (33%), Positives = 178/304 (57%), Gaps = 1/304 (0%)

Query: 17  VIGSLLSKGVSKENRDLANQQNITIGILQFVTHEALDDIKRGIEDQLK-KQMPQKQNVVI 75
           VI + +  G +  +      +   + + Q V H ALD  ++G+ D LK K   + +N+
Sbjct: 6   VIATAVLAGAALLSSQSIMAKTAKVAVSQIVEHPALDATRQGLLDGLKAKGYEEGKNLEF 65

Query: 76  KVMNAEGDQSKIQTMSRQLVQSGSDIVIGIATPAAQGLAATSKDIPVVMSAVSDPVGSRL 135
               A+G+ +    ++RQ V    D+++GIATP AQ L + +K IP+V +AV+DPVG++L
Sbjct: 66  DYKTAQGNPAIAVQIARQFVGENPDVLVGIATPTAQALVSATKTIPIVFTAVTDPVGAKL 125

Query: 136 VMQLDQPEANVTGLSNKVPVKQTIDLMKKLTPHVKTVGILYASNEDNSLSQVKEFRRLAR 195
```

```
           V QL+QP  NVTGLS+  PV+Q ++L+K++ P+VK++G++Y   E N++S ++  +  A
Sbjct: 126 VKQLEQPGKNVTGLSDLSPVEQHVELIKEILPNVKSIGVVYNPGEANAVSLMELLKLSAA 185

Query: 196 KKGYQVISYAVPSTNEVPATMSVMLGKVDAVFIPQDNTIASAFSSVMTTSKAAKIPVYTS 255
           K G +++      + +V +    +  K D ++   DNT+ASA   ++  +  AK PV+ +
Sbjct: 186 KHGIKLVEATALKSADVQSATQAIAEKSDVIYALIDNTVASAIEGMIVAANQAKTPVFGA 245

Query: 256 VDRMVEKGGLAAISQNQYDLGVQTANQVLKLIKGKRVVDVPVKVVDIGQPLINKNVAAEL 315
               VE+G +A++  + Y +GVQTA+ V  +++GK    + V+V      +INK  A +L
Sbjct: 246 ATSYVERGAIASLGFDYYQIGVQTADYVAAILEGKEPGSLDVQVAKGSDLVINKTAAEQL 305

Query: 316 GIAI 319
           GI I
Sbjct: 306 GITI 309
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 181/322 (56%), Positives = 252/322 (78%), Gaps = 1/322 (0%)

Query: 1   MKNKGLIATLILLTILVVGELFYNK-SEKRLNLSEKQVVKIGILQYVTHDALDAIEKGVE 59
           MKNK LIATL++LT++V+G L     S++  +L+ +Q + IGILQ+VTH+ALD I++G+E
Sbjct: 1   MKNKSLIATLLVLTVIVIGSLLSKGVSKENRDLANQQNITIGILQFVTHEALDDIKRGIE 60

Query: 60  DGLAQEGYKGKKVKLTVLNAEADQSKIQAMSKQLVNHHNDILIGIATPSAQGLAASTKDT 119
           D L ++  + + V + V+NAE DQSKIQ MS+QLV   +DI+IGIATP+AQGLAA++KD
Sbjct: 61  DQLKKQMPQKQNVVIKVMNAEGDQSKIQTMSRQLVQSGSDIVIGIATPAAQGLAATSKDI 120

Query: 120 PIIMGAVSDPLGAKLVTNMKKPTTNVTGLSNVVPTKQTVQLIKDITPNIKRIGILYASSE 179
           P++M AVSDP+G++LV  + +P  NVTGLSN VP KQT+ L+K +TP++K +GILYAS+E
Sbjct: 121 PVVMSAVSDPVGSRLVMQLDQPEANVTGLSNKVPVKQTIDLMKKLTPHVKTVGILYASNE 180

Query: 180 DNSVSQVTEFTKYAQKAGLEVLKYSVPSTNEIKTSMSVMTKKVDAVFVPQDNTIASAFRT 239
           DNS+SQV EF + A+K G +V+ Y+VPSTNE+  +MSVM  KVDAVF+PQDNTIASAF +
Sbjct: 181 DNSLSQVKEFRRLARKKGYQVISYAVPSTNEVPATMSVMLGKVDAVFIPQDNTIASAFSS 240

Query: 240 VIVAANQANIPVYSSVDTMVEQGSIASVAQSQYGLGLETAKQAIKVLRGKPVKDVPVKVI 299
           V+  +  A IPVY+SVD MVE+G +A+++Q+QY LG++TA Q +K+++GK V DVPVKV+
Sbjct: 241 VMTTSKAAKIPVYTSVDRMVEKGGLAAISQNQYDLGVQTANQVLKLIKGKRVVDVPVKVV 300

Query: 300 DTGKPSLNLKAAKHLGIKIPKK 321
           D G+P +N   A  LGI I K+
Sbjct: 301 DIGQPLINKNVAAELGIAIKKE 322
```

SEQ ID 2710 (GBS254) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 4; MW 27kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 3; MW 59.6kDa).

GBS254-GST was purified as shown in Figure 203, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 893

A DNA sequence (GBSx0947) was identified in *S.agalactiae* <SEQ ID 2713> which encodes the amino acid sequence <SEQ ID 2714>. This protein is predicted to be probable permease of ABC transporter (rbsC). Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-15.12    Transmembrane  127 - 143 ( 119 - 151)
   INTEGRAL    Likelihood = -8.81    Transmembrane  206 - 222 ( 200 - 227)
   INTEGRAL    Likelihood = -6.48    Transmembrane  260 - 276 ( 258 - 282)
   INTEGRAL    Likelihood = -5.84    Transmembrane  234 - 250 ( 231 - 257)
```

```
   INTEGRAL    Likelihood = -4.78   Transmembrane   55 -  71 (  54 -  72)
   INTEGRAL    Likelihood = -3.61   Transmembrane  177 - 193 ( 176 - 194)
   INTEGRAL    Likelihood = -3.35   Transmembrane   84 - 100 (  83 - 102)
   INTEGRAL    Likelihood = -1.91   Transmembrane   10 -  26 (  10 -  26)

----- Final Results -----
               bacterial membrane --- Certainty=0.7050(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG07224 GB:AE004801 probable permease of ABC transporter
          [Pseudomonas aeruginosa]
 Identities = 116/288 (40%), Positives = 185/288 (63%), Gaps = 9/288 (3%)

Query: 2   IISSVSQGLLWGILGLGIYLTFRILKFPDMTTEGSFPLGGAVCVTLMNQGVNPILATILG 61
           +  ++  GL++ ++ LG++++FR+L+FPD+T +GSFPLGGAVC TL+  G +P  AT+
Sbjct: 6   LFGALEIGLIFSLVALGVFISFRLLRFPDLTVDGSFPLGGAVCATLIALGWDPYSATLAA 65

Query: 62  MLSGMLAGFVTGLLYTKGKIPTILAGILVMTSCHSIMLMVMKRANLGLNEIQTLKDFLPF 121
             +G LAG  TGLL  K KI  +LA IL+M + +SI L +M + N+ L    TL   L
Sbjct: 66  TAAGALAGLATGLLNVKLKIMDLLASILMMIALYSINLRIMGKPNVPLIAEPTLFTLLQP 125

Query: 122 SNDLNLLVLGLIAILLVISA---LIYFLYTRLGQAYIATGDNPDMAKSFGIDTDKMEMLG 178
               + +   L+ + +VI+A   L +F  T+ G A  ATG NP MA++ G++T  M +LG
Sbjct: 126 EWLSDYVFRPLLLVFIVIAAKLLLDWFFTTQKGLAIRATGSNPRMARAQGVNTGGMILLG 185

Query: 179 LIVSNGLIALSGALVSQQDGYADVSKGIGVIVIGLASIIIGE-VLYSTGLTLFERLIAIV 237
           + +SN L+AL+GAL +Q  G AD+S GIG IVIGLA++I+GE +L S  L L    +A++
Sbjct: 186 MAISNALVALAGALFAQTQGGADISMGIGTIVIGLAAVIVGESILPSRRLIL--ATLAVI 243

Query: 238 VGSILYQFLITAVI---ALGFNTNYLKLFSAIVLGICLMVPVLKTKIL 282
           +G+I+Y+F I   +    +G     L L +A+++ + L++P++K ++L
Sbjct: 244 LGAIVYRFFIALALNSDFIGLQAQDLNLVTAVLVTVALVIPMMKKRLL 291
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2715> which encodes the amino acid sequence <SEQ ID 2716>. Analysis of this protein sequence reveals the following:

```
      Possible site: 55
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.46   Transmembrane  131 - 147 ( 125 - 156)
   INTEGRAL    Likelihood = -8.65   Transmembrane  210 - 226 ( 204 - 230)
   INTEGRAL    Likelihood = -8.17   Transmembrane  265 - 281 ( 261 - 283)
   INTEGRAL    Likelihood = -7.22   Transmembrane  238 - 254 ( 233 - 261)
   INTEGRAL    Likelihood = -3.03   Transmembrane   89 - 105 (  87 - 107)
   INTEGRAL    Likelihood = -2.60   Transmembrane   63 -  79 (  62 -  79)
   INTEGRAL    Likelihood = -2.23   Transmembrane  180 - 196 ( 180 - 198)
   INTEGRAL    Likelihood = -2.13   Transmembrane   14 -  30 (  14 -  30)

----- Final Results -----
               bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAG07224 GB:AE004801 probable permease of ABC transporter
          [Pseudomonas aeruginosa]
 Identities = 118/285 (41%), Positives = 186/285 (64%), Gaps = 7/285 (2%)

Query: 6   IISSVSQGLIWGVLGLGIYLTFRILNFPDMTTEGSFPLGGAVAVTAISLGWNPFLSTLLG 65
           +  ++  GLI+ ++ LG++++FR+L FPD+T +GSFPLGGAV  T I+LGW+P+ +TL
Sbjct: 6   LFGALEIGLIFSLVALGVFISFRLLRFPDLTVDGSFPLGGAVCATLIALGWDPYSATLAA 65

Query: 66  MLSGALAGFLTGLLYTKGKMPTLLAGILVMTSCNSIMLMVMGRANLGLHDHKRIQDCLPF 125
             +GALAG  TGLL  K K+  LLA IL+M +  SI L +MG+ N+ L     +   L
```

```
Sbjct: 66  TAAGALAGLATGLLNVKLKIMDLLASILMMIALYSINLRIMGKPNVPLIAEPTLFTLLQP 125

Query: 126 SIDLNSLLTGLITVVIVIS---VLIYFLYTNLGQAYIATGDNKDMAKSFGINTDWMEVMG 182
               + +   L+ V IVI+   +L +F  T  G A  ATG N  MA++ G+NT  M ++G
Sbjct: 126 EWLSDYVFRPLLLVFIVIAAKLLLDWFFTTQKGLAIRATGSNPRMARAQGVNTGGMILLG 185

Query: 183 LVVSNSLIALSGALVSQQDGYADVSKGIGVIVIGLASIIVGEVLYSTGLTLLERLIAIVI 242
           + +SN+L+AL+GAL +Q  G AD+S GIG IVIGLA++IVGE +  +   +L  L A+++
Sbjct: 186 MAISNALVALAGALFAQTQGGADISMGIGTIVIGLAAVIVGESILPSRRLILATL-AVIL 244

Query: 243 GSILYQFLISVVIT---LGFNTSYLKLISALVLALCLMIPVVKER 284
           G+I+Y+F I++ +    +G      L L++A+++ + L+IP++K+R
Sbjct: 245 GAIVYRFFIALALNSDFIGLQAQDLNLVTAVLVTVALVIPMMKKR 289
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 227/287 (79%), Positives = 259/287 (90%)

Query: 1   MIISSVSQGLLWGILGLGIYLTFRILKFPDMTTEGSFPLGGAVCVTLMNQGVNPILATIL 60
           MIISSVSQGL+WG+LGLGIYLTFRIL FPDMTTEGSFPLGGAV VT ++ G NP L+T+L
Sbjct: 5   MIISSVSQGLIWGVLGLGIYLTFRILNFPDMTTEGSFPLGGAVAVTAISLGWNPFLSTLL 64

Query: 61  GMLSGMLAGFVTGLLYTKGKIPTILAGILVMTSCHSIMLMVMKRANLGLNEIQTLKDFLP 120
           GMLSG LAGF+TGLLYTKGK+PT+LAGILVMTSC+SIMLMVM RANLGL++ + ++D LP
Sbjct: 65  GMLSGALAGFLTGLLYTKGKMPTLLAGILVMTSCNSIMLMVMGRANLGLHDHKRIQDCLP 124

Query: 121 FSNDLNLLVLGLIAILLVISALIYFLYTRLGQAYIATGDNPDMAKSFGIDTDKMEMLGLI 180
           FS DLN L+ GLI +++VIS LIYFLYT LGQAYIATGDN DMAKSFGI+TD ME++GL+
Sbjct: 125 FSIDLNSLLTGLITVVIVISVLIYFLYTNLGQAYIATGDNKDMAKSFGINTDWMEVMGLV 184

Query: 181 VSNGLIALSGALVSQQDGYADVSKGIGVIVIGLASIIIGEVLYSTGLTLFERLIAIVVGS 240
           VSN LIALSGALVSQQDGYADVSKGIGVIVIGLASII+GEVLYSTGLTL ERLIAIV+GS
Sbjct: 185 VSNSLIALSGALVSQQDGYADVSKGIGVIVIGLASIIVGEVLYSTGLTLLERLIAIVIGS 244

Query: 241 ILYQFLITAVIALGFNTNYLKLFSAIVLGICLMVPVLKTKILKGVRL 287
           ILYQFLI+ VI LGFNT+YLKL SA+VL +CLM+PV+K +  KGVRL
Sbjct: 245 ILYQFLISVVITLGFNTSYLKLISALVLALCLMIPVVKERFFKGVRL 291
```

A related GBS gene <SEQ ID 8681> and protein <SEQ ID 8682> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
McG: Discrim Score:      4.24
GvH: Signal Score (-7.5): -6.43
     Possible site: 24
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 8 value: -15.12 threshold:  0.0
   INTEGRAL    Likelihood =-15.12   Transmembrane  127 - 143 ( 119 - 151)
   INTEGRAL    Likelihood = -7.54   Transmembrane  206 - 222 ( 201 - 225)
   INTEGRAL    Likelihood = -6.48   Transmembrane  260 - 276 ( 258 - 282)
   INTEGRAL    Likelihood = -5.84   Transmembrane  234 - 250 ( 231 - 257)
   INTEGRAL    Likelihood = -4.78   Transmembrane   55 -  71 (  54 -  72)
   INTEGRAL    Likelihood = -3.61   Transmembrane  177 - 193 ( 176 - 194)
   INTEGRAL    Likelihood = -3.35   Transmembrane   84 - 100 (  83 - 102)
   INTEGRAL    Likelihood = -1.91   Transmembrane   10 -  26 (  10 -  26)
   PERIPHERAL  Likelihood =  4.77       36
 modified ALOM score:   3.52

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.7050(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00338(298 - 1146 of 1461)
GP|9950013|gb|AAG07224.1|AE004801_2|AE004801(4 - 291 of 296) probable permease of ABC
transporter {Pseudomonas aeruginosa}
%Match = 20.2
%Identity = 40.8  %Similarity = 68.3
Matches = 116  Mismatches = 84  Conservative Sub.s = 78

126       156       186       216       246       276       306       336
YGLGLETAKQAIKVLRGKPVKDVPVKVIDTGKPSLNLKAAKHLGIKIPKKIMKQAEITVKVDD*KEGFMIISSVSQGLLW
                                                                   | :  ::  ||::
                                                                MSLFSLFGALEIGLIF
                                                                        10

366       396       426       456       486       516       546       576
GILGLGIYLTFRILKFPDMTTEGSFPLGGAVCVTLMNQGVNPILATILGMLSGMLAGFVTGLLYTKGKIPTILAGILVMT
 :: ||::::||:|:|||:| :|||||||||| ||:  | :|  ||:    :| |||: ||||  | ||  :|| ||:|
SLVALGVFISFRLLRFPDLTVDGSFPLGGAVCATLIALGWDPYSATLAATAAGALAGLATGLLNVKLKIMDLLASILMMI
            30        40        50        60        70        80        90

606       636                 690       720       747       777       807
SCHSIMLMVMKRANLGLNEIQTLKDFL-P-FSNDLNLLVLGLIAILLVISALI-YFLYTRLGQAYIATGDNPDMAKSFGI
: :|| | :| : |: |     ||  :| | : :|  :  | |: |::    |: :|: |: | |  ||| || ||:: |:
ALYSINLRIMGKPNVPLIAEPTLFTLLQPEWLSDYVFRPLLLVFIVIAAKLLLDWFFTTQKGLAIRATGSNPRMARAQGV
           110       120       130       140       150       160       170

837       867       897       927       957       987       1017      1047
DTDKMEMLGLIVSNGLIALSGALVSQQDGYADVSKGIGVIXIGLASIIIGEVLYSTGLTLFERLIAIVVGSILYQFLITA
:|  | :||: :|| |:||:||| :|  | ||:| ||| | ||||::|:|| :  :   ::  | |:::|:|:|:|:|
NTGGMILLGMAISNALVALAGALFAQTQGGADISMGIGTIVIGLAAVIVGESILPSRRLILATL-AVILGAIVYRFFI--
           190       200       210       220       230       240        250

1077      1086      1116      1146      1176      1206      1236      1266
VIALGFNTNY-------LKLFSAIVLGICLMVPVLKTKILKGVRL*W**KS*S*KKQPYKSVMV*QK*KRY*IMLI*VFM
  || :|:::       | | :|::: : |::|::| ::|
--ALALNSDFIGLQAQDLNLVTAVLVTVALVIPMMKKRLLGKKGA
                270       280       290
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 894

A DNA sequence (GBSx0948) was identified in *S.agalactiae* <SEQ ID 2717> which encodes the amino acid sequence <SEQ ID 2718>. This protein is predicted to be ABC transporter (potA). Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9887> which encodes amino acid sequence <SEQ ID 9888> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF86640 GB:AF162694 ABC transporter [Enterococcus gallinarum]
 Identities = 117/252 (46%), Positives = 167/252 (65%)

Query: 19  MVMKIIELKEATVQVSNGLAEMKTILDHVNLSIYEHDFITILGGNGAGKSTLFNVIAGTL 78
            M   ++ + +       G      +L  ++L++   DFITI+GGNGAGKSTL N IAGT+
```

```
Sbjct: 1   MTTPVLTISDLHQTFEKGTINENHVLRGIDLTMNSGDFITIIGGNGAGKSTLLNSIAGTI 60

Query: 79  MLSSGNIYIMGQDVTNLSAEKRAKYLSRVFQDPKMGTAPRMTVAENLLVAKFRGEKRPLV 138
                G I +  +++T  S  +R+K +SRVFQDP+MGTA R+TV ENL +A  RG+ R
Sbjct: 61  PTEQGKIVLGDKEITRHSVTRRSKEISRVFQDPRMGTAVRLTVEENLALAYKRGQVRGFS 120

Query: 139 PRKIINYTEEFQKLIARTGNGLDRHLETPTGLLSGGQRQALSLLMATLKKPNLLLLDEHT 198
                  +   F++ +AR   GL+  L T  GLLSGGQRQA++LLMATL++P L+LLDEHT
Sbjct: 121 SGVKGKHRAFFKEKLARLNLGLENRLTTEIGLLSGGQRQAITLLMATLQQPKLILLDEHT 180

Query: 199 AALDPRTSVSLMGLTDEFIKQDSLTALMITHHMEDALKYGNRVLVMKDGKIVRDLNQAQK 258
           AALDP+TS+++M LTD+ I++  LTA M+TH MEDA++YGNR++++  GKIV D+   +K
Sbjct: 181 AALDPKTSMTVMALTDQLIQEQQLTAFMVTHDMEDAIRYGNRLIMLHQGKIVVDITGEEK 240

Query: 259 NKMAIADYYQLF 270
             + + D   LF
Sbjct: 241 QSLTVPDLMALF 252
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2719> which encodes the amino acid sequence <SEQ ID 2720>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2249(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities .= 186/250 (74%), Positives = 210/250 (83%)

Query: 22  KIIELKEATVQVSNGLAEMKTILDHVNLSIYEHDFITILGGNGAGKSTLFNVIAGTLMLS 81
           KIIEL  ATV V NG  + KTILD+V L+IYEHDF+TILGGNGAGKSTLFNVIAGTL L+
Sbjct: 3   KIIELINATVDVDNGFEDAKTILDNVTLTIYEHDFLTILGGNGAGKSTLFNVIAGTLSLT 62

Query: 82  SGNIYIMGQDVTNLSAEKRAKYLSRVFQDPKMGTAPRMTVAENLLVAKFRGEKRPLVPRK 141
            G I I+GQDVT+  AEKRA YLSRVFQD KMGTAPRMTVAENLL+A+ RG KR L  RK
Sbjct: 63  RGQIRILGQDVTHWPAEKRALYLSRVFQDSKMGTAPRMTVAENLLIARQRGGKRSLASRK 122

Query: 142 IINYTEEFQKLIARTGNGLDRHLETPTGLLSGGQRQALSLLMATLKKPNLLLLDEHTAAL 201
           I  +   F+ L+ RTGNGL++HLETP GLLSGGQRQALSLLMATLKKP LLLLDEHTAAL
Sbjct: 123 ITEHLASFEDLVKRTGNGLEKHLETPAGLLSGGQRQALSLLMATLKKPALLLLDEHTAAL 182

Query: 202 DPRTSVSLMGLTDEFIKQDSLTALMITHHMEDALKYGNRVLVMKDGKIVRDLNQAQKNKM 261
           DP+TS SLM LTDEF+ +D LTALMITHHMEDAL YGNR++VMKDG I++DLNQ +K ++
Sbjct: 183 DPKTSQSLMQLTDEFVTKDGLTALMITHHMEDALTYGNRLIVMKDGNIIKDLNQMEKEQL 242

Query: 262 AIADYYQLFD 271
            I DYYQLFD
Sbjct: 243 TITDYYQLFD 252
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 895

A DNA sequence (GBSx0949) was identified in *S.agalactiae* <SEQ ID 2721> which encodes the amino acid sequence <SEQ ID 2722>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
               bacterial cytoplasm --- Certainty=0.1930(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
RGD motif: 415-417
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAB06117 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 236/549 (42%), Positives = 362/549 (64%), Gaps = 2/549 (0%)

Query: 4   IKIMALGGVRENGKNLYVVEVNDSIFVLDAGLKYPENEQLGVDVVIPNLDYLIENKKRVQ 63
           I++ ALGGV E GKN+YVVEV+D +FV+DAGL +P++E LGVDVVIP++ YL+EN++RV+
Sbjct: 9   IRVFALGGVGEIGKNMYVVEVDDDLFVIDAGLMFPDDEMLGVDVVIPDISYLVENEERVR 68

Query: 64  GIFLTHGHADAIGALPYIIAEVKAPVFGSPLTIELAKLFVKNSTAVKKFNNFHVIDSETE 123
            I LTHGH D IG LPY++ ++  PV+G+ LT+ L +  +K +  ++      +IDS +
Sbjct: 69  AILLTHGHEDHIGGLPYVLQKLNVPVYGTKLTLGLVEEKLKEAGLIRSAK-LKLIDSNSR 127

Query: 124 IEFQDAVISFFKTTHSIPESMGIVIGTKEGNIVYTGDFKFDQAARKYYQTDLARLAEIGR 183
           ++     +SFF+T HSIP+S+GI I T +G IV+TGDFKFDQ      Q ++ ++A IG
Sbjct: 128 LKLGSTPVSFFRTNHSIPDSVGICIQTSQGFIVHTGDFKFDQTPVDGKQAEIGKMAAIGH 187

Query: 184 DGVLALLSDSANATSNEQVASEYEVGDEIKSVIEDAEGRVIVAAVASNLIRIQQVFDAAA 243
            GVL LLSDS NA      SE EVG  I    E  +GR+IV   ASN+ R+QQV  AA
Sbjct: 188 KGVLCLLSDSTNAERPGMTKSETEVGRGIAEAFEQTKGRIIVTTFASNVHRVQQVIHAAI 247

Query: 244 ENGRRVVLTGFDIENIVRTAIRMKRIHIADENMIIKPKDMTRYEDNELLILETGRMGEPI 303
              R++ + G  +  +V  A R+  +   D+ + I  +++++Y+D  + I+ TG  GEP+
Sbjct: 248 ATNRKLAVAGRSMVKVVSIAERLGYLEAPDD-LFIDIEEVSKYDDERVAIITTGSQGEPM 306

Query: 304 NGLQKMAIGRHRYVQIKDGDLVFIVTTPSIAKEAVVARVENLIYKAGGSVKLITQNLRVS 363
           + L +MA G HR + I + D V I  TP    E  V+ + +L+++ G  V      +  S
Sbjct: 307 SALSRMAKGAHRQITITENDTVIIAATPIPGNERSVSTIVDLLHRIGADVIFGHGKVHAS 366

Query: 364 GHANGRELQLLMNLLKPKYLFPIQGEYRDLSAHAGLAQEVGMSADDIYIVKRGDIMVLEK 423
           GH +  EL+L++NL++PK+  PI GE+R   AH  LA+ VG+  + I++V +G+++
Sbjct: 367 GHGSAEELKLMLNLMRPKFFVPIHGEFRMQHAHKELAKSVGIREEAIFLVDKGEVVEFRN 426

Query: 424 DGFFHSGSVPAGDVMIDGNAIGDVGNIVLRDRKVLSEDGIFIVVITVSKKEKKIISKARV 483
                +G VP+G+V+IDG  +GDVGNIVLRDR++LS+DGI +VV+T++K+    I+S   +
Sbjct: 427 GQGRKAGKVPSGNVLIDGLGVGDVGNIVLRDRRLLSKDGILVVVVTLNKQSGTILSGPNI 486

Query: 484 NTRGFVYVKKSRDILRESAELVNTTVEDYLSKDTFDWGELKGKVRDEVSKFLFDQTKRRP 543
            +RGFVYV++S  ++ E+ ELV  T++  ++++  +W  LK  VR+ +S+FLF++TKRRP
Sbjct: 487 ISRGFVYVRESEKLIEEANELVTETLKKCVTENVNEWSSLKSNVREVLSRFLFEKTKRRP 546

Query: 544 AILPVVMEV 552
            ILP++MEV
Sbjct: 547 MILPIIMEV 555
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2723> which encodes the amino acid sequence <SEQ ID 2724>. Analysis of this protein sequence reveals the following:

```
        Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2204(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB06117 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 232/549 (42%), Positives = 360/549 (65%), Gaps = 2/549 (0%)
```

```
Query: 4   IKMIALGGVREYGKNFYLVEINDSMFILDAGLKYPENEQLGVDLVIPNLDYVIENKGKVQ 63
           I++ ALGGV E GKN Y+VE++D +F++DAGL +P++E LGVD+VIP++ Y++EN+ +V+
Sbjct: 9   IRVFALGGVGEIGKNMYVVEVDDDLFVIDAGLMFPDDEMLGVDVVIPDISYLVENEERVR 68

Query: 64  GIFLSHGHADAIGALPYLLAEVSAPVFGSELTIELAKLFVKSNNSTKKFNNFHVVDSDTE 123
            I L+HGH D IG LPY+L +++ PV+G++LT+ L +  +K     +      ++DS++
Sbjct: 69  AILLTHGHEDHIGGLPYVLQKLNVPVYGTKLTLGLVEEKLKEAGLIRSAK-LKLIDSNSR 127

Query: 124 IEFKDGLVSFFRTTHSIPESMGIVIGTDKGNIIYTGDFKFDQAAREGYQTDLLRLAEIGK 183
           ++     VSFFRT HSIP+S+GI I T +G I++TGDFKFDQ   +G Q ++ ++A IG
Sbjct: 128 LKLGSTPVSFFRTNHSIPDSVGICIQTSQGFIVHTGDFKFDQTPVDGKQAEIGKMAAIGH 187

Query: 184 EGVLALLSDSVNATSNDQIASESEVGEEMDSVISDADGRVIVAAVASNLVRIQQVFDSAT 243
           +GVL LLSDS NA       SE+EVG  +        GR+IV   ASN+ R+QQV  +A
Sbjct: 188 KGVLCLLSDSTNAERPGMTKSETEVGRGIAEAFEQTKGRIIVTTFASNVHRVQQVIHAAI 247

Query: 244 AHGRRVVLTGTDAENIVRTALRLEKLMITDERLLIKPKDMSKFEDHELIILEAGRMGEPI 303
           A  R++ + G     +V  A RL  L   D+ L I  +++SK++D  + I+  G  GEP+
Sbjct: 248 ATNRKLAVAGRSMVKVVSIAERLGYLEAPDD-LFIDIEEVSKYDDERVAIITTGSQGEPM 306

Query: 304 NSLQKMAAGRHRYVQIKEGDLVYIVTTPSTAKEAMVARVENLIYKAGGSVKLITQNLRVS 363
           ++L +MA G HR + I E D V I  TP    E  V+ + +L+++ G  V       + S
Sbjct: 307 SALSRMAKGAHRQITITENDTVIIAATPIPGNERSVSTIVDLLHRIGADVIFGHGKVHAS 366

Query: 364 GHANGRDLQLLMNLLKPQYLFPVQGEYRDLAAHAKLAEEVGIFPENIHILKRGDIMVLND 423
           GH +  +L+L++NL++P++  P+ GE+R   AH +LA+ VGI  E I ++ +G+++   +
Sbjct: 367 GHGSAEELKLMLNLMRPKFFVPIHGEFRMQHAHKELAKSVGIREEAIFLVDKGEVVEFRN 426

Query: 424 EGFLHEGGVPASDVMIDGNAIGDVGNIVLRDRKVLSEDGIFIVAITVSKKEKRIISKAKV 483
                 G VP+ +V+IDG  +GDVGNIVLRDR++LS+DGI +V +T++K+   I+S   +
Sbjct: 427 GQGRKAGKVPSGNVLIDGLGVGDVGNIVLRDRRLLSKDGILVVVVTLNKQSGTILSGPNI 486

Query: 484 NTRGFVYVKKSHDILRESAELVNTTVGNYLKKDTFDWGELKGNVRDDLSKFLFEQTKRRP 543
            +RGFVYV++S  ++ E+ ELV  T+   + ++  +W  LK NVR+ LS+FLFE+TKRRP
Sbjct: 487 ISRGFVYVRESEKLIEEANELVTETLKKCVTENVNEWSSLKSNVREVLSRFLFEKTKRRP 546

Query: 544 AILPVVMEV 552
            ILP++MEV
Sbjct: 547 MILPIIMEV 555
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 446/553 (80%), Positives = 513/553 (92%)

Query: 1   MSDIKIMALGGVRENGKNLYVVEVNDSIFVLDAGLKYPENEQLGVDVVIPNLDYLIENKK 60
           M+DIK++ALGGVRE GKN Y+VE+NDS+F+LDAGLKYPENEQLGVD+VIPNLDY+IENK
Sbjct: 1   MTDIKMIALGGVREYGKNFYLVEINDSMFILDAGLKYPENEQLGVDLVIPNLDYVIENKG 60

Query: 61  RVQGIFLTHGHADAIGALPYIIAEVKAPVFGSPLTIELAKLFVKNSTAVKKFNNFHVIDS 120
           +VQGIFL+HGHADAIGALPY++AEV APVFGS LTIELAKLFVK++ + KKFNNFHV+DS
Sbjct: 61  KVQGIFLSHGHADAIGALPYLLAEVSAPVFGSELTIELAKLFVKSNNSTKKFNNFHVVDS 120

Query: 121 ETEIEFQDAVISFFKTTHSIPESMGIVIGTKEGNIVYTGDFKFDQAARKYYQTDLARLAE 180
           +TEIEF+D ++SFF+TTHSIPESMGIVIGT +GNI+YTGDFKFDQAAR+ YQTDL RLAE
Sbjct: 121 DTEIEFKDGLVSFFRTTHSIPESMGIVIGTDKGNIIYTGDFKFDQAAREGYQTDLLRLAE 180

Query: 181 IGRDGVLALLSDSANATSNEQVASEYEVGDEIKSVIEDAEGRVIVAAVASNLIRIQQVFD 240
           IG++GVLALLSDS NATSN+Q+ASE EVG+E+ SVI DA+GRVIVAAVASNL+RIQQVFD
Sbjct: 181 IGKEGVLALLSDSVNATSNDQIASESEVGEEMDSVISDADGRVIVAAVASNLVRIQQVFD 240

Query: 241 AAAENGRRVVLTGFDIENIVRTAIRMKRIHIADENMIIKPKDMTRYEDNELLILETGRMG 300
           +A  +GRRVVLTG D ENIVRTA+R++++ I DE ++IKPKDM+++ED+EL+ILE GRMG
Sbjct: 241 SATAHGRRVVLTGTDAENIVRTALRLEKLMITDERLLIKPKDMSKFEDHELIILEAGRMG 300

Query: 301 EPINGLQKMAIGRHRYVQIKDGDLVFIVTTPSIAKEAVVARVENLIYKAGGSVKLITQNL 360
           EPIN LQKMA GRHRYVQIK+GDLV+IVTTPS AKEA+VARVENLIYKAGGSVKLITQNL
Sbjct: 301 EPINSLQKMAAGRHRYVQIKEGDLVYIVTTPSTAKEAMVARVENLIYKAGGSVKLITQNL 360
```

```
Query: 361 RVSGHANGRELQLLMNLLKPKYLFPIQGEYRDLSAHAGLAQEVGMSADDIYIVKRGDIMV 420
           RVSGHANGR+LQLLMNLLKP+YLFP+QGEYRDL+AHA LA+EVG+  ++I+I+KRGDIMV
Sbjct: 361 RVSGHANGRDLQLLMNLLKPQYLFPVQGEYRDLAAHAKLAEEVGIFPENIHILKRGDIMV 420

Query: 421 LEKDGFFHSGSVPAGDVMIDGNAIGDVGNIVLRDRKVLSEDGIFIVVITVSKKEKKIISK 480
           L  +GF H G VPA DVMIDGNAIGDVGNIVLRDRKVLSEDGIFIV ITVSKKEK+IISK
Sbjct: 421 LNDEGFLHEGGVPASDVMIDGNAIGDVGNIVLRDRKVLSEDGIFIVAITVSKKEKRIISK 480

Query: 481 ARVNTRGFVYVKKSRDILRESAELVNTTVEDYLSKDTFDWGELKGKVRDEVSKFLFDQTK 540
           A+VNTRGFVYVKKS DILRESAELVNTTV +YL KDTFDWGELKG VRD++SKFLF+QTK
Sbjct: 481 AKVNTRGFVYVKKSHDILRESAELVNTTVGNYLKKDTFDWGELKGNVRDDLSKFLFEQTK 540

Query: 541 RRPAILPVVMEVR 553
           RRPAILPVVMEVR
Sbjct: 541 RRPAILPVVMEVR 553
```

There is also homology to SEQ ID 4910.

SEQ ID 2722 (GBS295) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 2; MW 89.4kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 167 (lane 9 & 11; MW 79kDa – thioredoxin fusion) and in Figure 238 (lane 3; MW 79kDa – thioredoxin fusion).

Purified Thio-GBS295-His is shown in Figure 244, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 896

A DNA sequence (GBSx0950) was identified in *S.agalactiae* <SEQ ID 2725> which encodes the amino acid sequence <SEQ ID 2726>. This protein is predicted to be tributyrin esterase. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
          bacterial outside --- Certainty=0.3000(Affirmative) < succ>
         bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9885> which encodes amino acid sequence <SEQ ID 9886> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF62859 GB:AF157484 tributyrin esterase [Lactococcus lactis
           subsp. lactis]
 Identities = 154/262 (58%), Positives = 188/262 (70%), Gaps = 4/262 (1%)

Query: 21  MAFFNIEYHSKVLGTERQVNVIYPDAFEMSDDKIDDCDIPVLYLLHGMGGNENSWQKRTN 80
           MA  NIEY+S+VLG  R+VNVIYP++ ++ D      DIPVLYLLHGM GNENSW  R+
Sbjct: 1   MAVINIEYYSEVLGMNRKVNVIYPESSKVED--FTQTDIPVLYLLHGMSGNENSWIIRSG 58

Query: 81  IERLLRHTNLIVVMPSTDLAWYTNTKYGLDYFDAIAIELPKVLKRFFPNMSDKREKNFIA 140
           IERL+RHTNL +VMPSTDL +Y NT YG++YFDAIA ELPKV+  FFPN+S KREKNFIA
Sbjct: 59  IERLIRHTNLAIVMPSTDLGFYVNTTYGMNYFDAIAHELPKVINNFFPNLSTKREKNFIA 118

Query: 141 GLSMGGYGAYKIALLTNRFSHAASLSGALSFDFDLLFNNGNNNINYWSGIFGDLNNTDNI 200
           GLSMGGYGAY++AL T+ FS+AASLSG L+FD   +  N   N  YW GIFG+
```

```
Sbjct: 119 GLSMGGYGAYRLALGTDYFSYAASLSGVLTFDG--MEENFKENPAYWGGIFGNWETFKGS 176

Query: 201 ERHSLRRYVESFDMKTKFYAWCGYEDFLFEANEVAIDELRQLGLTIDYFNDHGKHEWYYW 260
           +   L       + K K YAWCG +DFLF  NE A  EL++LG  I Y +  G HEWYYW
Sbjct: 177 DNEILSLADRKQENKPKLYAWCGKQDFLFPGNEYATAELKKLGFDITYESSDGVHEWYYW 236

Query: 261 NQQLEKVLEWLPVDYVKEERLS 282
            Q++E VL+WLP++Y +EERLS
Sbjct: 237 TQKIESVLKWLPINYKQEERLS 258
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2727> which encodes the amino acid sequence <SEQ ID 2728>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2183(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 172/262 (65%), Positives = 199/262 (75%), Gaps = 1/262 (0%)

Query: 21  MAFFNIEYHSKVLGTERQVNVIYPDAFEMSDDKIDDCDIPVLYLLHGMGGNENSWQKRTN 80
           MA   IEYHS VLG ER+VNVIYPD  E+      D DIPVLYLLHGMGGNENSWQKRT
Sbjct: 1   MASIAIEYHSVVLGMERKVNVIYPDQSEIPKKDQGDKDIPVLYLLHGMGGNENSWQKRTA 60

Query: 81  IERLLRHTNLIVVMPSTDLAWYTNTKYGLDYFDAIAIELPKVLKRFFPNMSDKREKNFIA 140
           IERLLRHTNLIVVMPSTDL WYT+T YGL+Y+ A++ ELP+VL  FFPNM+ KREK F+A
Sbjct: 61  IERLLRHTNLIVVMPSTDLGWYTDTAYGLNYYRALSQELPQVLAAFFPNMTQKREKTFVA 120

Query: 141 GLSMGGYGAYKIALLTNRFSHAASLSGALSFDFDLLFNNGNNNINYWSGIFGDLNNTDNI 200
           GLSMGGYGA+K AL +NRFS+AAS SGAL F  + L       + YW G+FG  ++ D +
Sbjct: 121 GLSMGGYGAFKWALKSNRFSYAASFSGALDFSPETLLEGKLGELAYWQGVFGQFDDPD-L 179

Query: 201 ERHSLRRYVESFDMKTKFYAWCGYEDFLFEANEVAIDELRQLGLTIDYFNDHGKHEWYYW 260
           ++H L+  V   D KTKFYAWCGYEDFLF  NE AI + +  GL IDY   HGKHEWYYW
Sbjct: 180 DKHYLKNMVAESDGKTKFYAWCGYEDFLFATNEKAIADFQAQGLDIDYHKGHGKHEWYYW 239

Query: 261 NQQLEKVLEWLPVDYVKEERLS 282
           NQQLE +LEWLP++Y KEERLS
Sbjct: 240 NQQLEVLLEWLPINYQKEERLS 261
```

SEQ ID 2726 (GBS645) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 129 (lanes 8 & 10; MW 60kDa + lane 9; MW 27kDa) and in Figure 186 (lane 4; MW 60kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 129 (lane 12; MW 34.7kDa), in Figure 140 (lane 8; MW 35kDa) and in Figure 178 (lane 4; MW 35kDa). Purified GBS645-GST is shown in Figure 236, lane 11; purified GBS645-His is shown in Figure 229, lanes 3-4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 897

A DNA sequence (GBSx0951) was identified in *S.agalactiae* <SEQ ID 2729> which encodes the amino acid sequence <SEQ ID 2730>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
```

```
   INTEGRAL    Likelihood = -9.34   Transmembrane   22 -  38 (  18 -  46)

----- Final Results -----
              bacterial membrane --- Certainty=0.4736(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2731> which encodes the amino acid sequence <SEQ ID 2732>. Analysis of this protein sequence reveals the following:

```
     Possible site: 52

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -7.43   Transmembrane   25 -  41 (  20 -  46)
   INTEGRAL    Likelihood = -2.71   Transmembrane    4 -  20 (   3 -  20)

----- Final Results -----
              bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 31/87 (35%), Positives = 50/87 (56%), Gaps = 2/87 (2%)

Query: 1  MRTLFRMIFAIPKFIFRLIWNIIWGIFKTVLVIAIILFGLYYYANHSQSEFANQLSDIIQ 60
           M+ L  +I  +PK I ++ W++I G  +T+L++ II+ GL YY+NHS S  AN++S  I
Sbjct: 1  MKQLLAIILWLPKLIVKMFWHLIKGFLQTILLVTIIIIGLMYYSNHSDSVLANKIS--IV 58

Query: 61 TGKTFLNFADTNQLKNSFTNLATDNVH 87
           T +    F    Q  ++ T   + N H
Sbjct: 59 TEQVVQIFDILTQKPSAKTRHGSGNSH 85
```

SEQ ID 2730 (GBS220d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 11-13; MW 50kDa) and in Figure 239 (lane 12; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 14-16; MW 25.2kDa) and in Figure 184 (lane 7; MW 25kDa). Purified GBS220d-GST is shown in Figure 246, lanes 3 & 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 898

A DNA sequence (GBSx0953) was identified in *S.agalactiae* <SEQ ID 2733> which encodes the amino acid sequence <SEQ ID 2734>. This protein is predicted to be unnamed protein product (rpiA). Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2538(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB69583 GB:A93589 unnamed protein product [Spinacia oleracea]
 Identities = 114/232 (49%), Positives = 147/232 (63%), Gaps = 11/232 (4%)

Query: 2   DELKKLAGVTAAKYVKNGMIVGLGTGSTAYFFVEEIGRRVKEEGL-QVVGVTTSNRTTEQ 60
            D+LKKLA   A   VK+GM++GLGTGSTA F V  IG  +    L  +VG+ TS RT EQ
Sbjct: 59  DDLKKLAAEKAVDSVKSGMVLGLGTGSTAAFAVSRIGELLSAGKLTNIVGIPTSKRTAEQ 118

Query: 61  ARGLGIPLKSADDIDVIDVTVDGADEVDPDFNGIKGGGGALLMEKIVATPTKEYIWVVDE 120
            A  LGIPL   DD   ID+ +DGADEVDPD N +KG GGALL EK+V   + ++I VVD+
Sbjct: 119 AASLGIPLSVLDDHPRIDLAIDGADEVDPDLNLVKGRGGALLREKMVEAASDKFIVVVDD 178

Query: 121 SKLVETLGAFKL--PVEVV----RYGSERLFRVFKSKGYCPSFRETEGDR--FITDMGNY 172
            +KLV+ LG  +L  PVEVV    +Y  +RL  +FK  G C +    EGD   ++TD  NY
Sbjct: 179 TKLVDGLGGSRLAMPVEVVQFCWKYNLKRLQEIFKELG-CEAKLRMEGDSSPYVTDNSNY 237

Query: 173 IIDLDL-KKIEDPKQLANELDHTVGVVEHGLFNGMVNKVIVAGKNGLDILEK 223
            I+DL     I+D +    E+    GVVEHGLF GM ++VI+AGK G+ +  K
Sbjct: 238 IVDLYFPTSIKDAEAAGREISALEGVVEHGLFLGMASEVIIAGKTGVSVKTK 289
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2735> which encodes the amino acid sequence <SEQ ID 2736>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1646(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 166/222 (74%), Positives = 190/222 (84%)

Query: 1   MDELKKLAGVTAAKYVKNGMIVGLGTGSTAYFFVEEIGRRVKEEGLQVVGVTTSNRTTEQ 60
            M+ LKK+AGVTAA+YV +GM +GLGTGSTAY+FVEEIGRRVK+EGLQVVGVTTS+ T++Q
Sbjct: 1   MEALKKIAGVTAAQYVTDGMTIGLGTGSTAYYFVEEIGRRVKQEGLQVVGVTTSSVTSKQ 60

Query: 61  ARGLGIPLKSADDIDVIDVTVDGADEVDPDFNGIKGGGGALLMEKIVATPTKEYIWVVDE 120
            A  LGIPLKS DDID ID+TVDGADEVD +FNGIKGGG ALLMEKIVATPTKEYIWVVD
Sbjct: 61  AEVLGIPLKSIDDIDSIDLTVDGADEVDKNFNGIKGGGAALLMEKIVATPTKEYIWVVDA 120

Query: 121 SKLVETLGAFKLPVEVVRYGSERLFRVFKSKGYCPSFRETEGDRFITDMGNYIIDLDLKK 180
            SK+VE LGAFKLPVEVV+YG++RLFRVF+  GY PSFR     R +TDM NYIIDLDL
Sbjct: 121 SKMVEHLGAFKLPVEVVQYGADRLFRVFEKAGYKPSFRMKGDSRLVTDMQNYIIDLDLGC 180

Query: 181 IEDPKQLANELDHTVGVVEHGLFNGMVNKVIVAGKNGLDILE 222
            I+DP    + LD TVGVVEHGLFNGMV+KVIVA K+G+ +LE
Sbjct: 181 IKDPVAFGHLLDGTVGVVEHGLFNGMVDKVIVASKDGVTVLE 222
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 899

A DNA sequence (GBSx0954) was identified in *S.agalactiae* <SEQ ID 2737> which encodes the amino acid sequence <SEQ ID 2738>. This protein is predicted to be phosphopentomutase (deoB). Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0546(Affirmative) < succ>
```

```
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC45496 GB:U80410 phosphopentomutase [Lactococcus lactis subsp. cremoris]
 Identities = 275/408 (67%), Positives = 325/408 (79%), Gaps = 7/408 (1%)

Query: 3   QFDRIHLVVLDSVGIGAAPDANDFVNAGVP------DGASDTLGHISKTVGLAVPNMAKI 56
            +F RIHLVV+DSVGIGAAPDA+ F N  V       D  SDT+GHIS+  GL VPN+ K+
Sbjct: 4   KFGRIHLVVMDSVGIGAAPDADKFFNHDVETHEAINDVKSDTIGHISEIRGLDVPNLQKL 63

Query: 57  GLGNIPRPQALKTVPAEENPSGYATKLQEVSLGKDTMTGHWEIMGLNITEPFDTFWNGFP 116
            G GNIPR   LKT+PA + P+ Y TKL+E+S GKDTMTGHWEIMGLNI  PF T+  G+P
Sbjct: 64  GWGNIPRESPLKTIPAAQKPAAYVTKLEEISKGKDTMTGHWEIMGLNIQTPFPTYPEGYP 123

Query: 117 EDIITKIEDFSGRKVIREANKPYSGTAVIDDFGPRQMETGELIIYTSADPVLQIAAHEDI 176
            ED++ KIE+FSGRK+IREANKPYSGTAVI+DFGPRQ+ETGELIIYTSADPVLQIAAHED+
Sbjct: 124 EDLLEKIEEFSGRKIIREANKPYSGTAVIEDFGPRQLETGELIIYTSADPVLQIAAHEDV 183

Query: 177 IPLEELYRICEYARSITMERPALL-GRIIARPYVGEPGNFTRTANRHDYAVSPFEDTVLN 235
            I  EELY+ICEY RSIT+E   ++ GRIIARPYVGE GNF RT  R DYA+SPF +TVL
Sbjct: 184 ISREELYKICEYVRSITLEGSGIMIGRIIARPYVGEAGNFERTDGRRDYALSPFAETVLE 243

Query: 236 KLDQAGIDTYAVGKINDIFNGSGINHDMGHNKSNSHGIDTLIKTMGLSEFEKGFSFTNLV 295
            KL +AGIDTY+VGKI+DIFN  G+ +DMGHN ++  G+D L+K M  +EF +GFSFTNLV
Sbjct: 244 KLYKAGIDTYSVGKISDIFNTVGVKYDMGHNHNDMDGVDRLLKAMTKTEFTEGFSFTNLV 303

Query: 296 DFDALYGHRRDPHGYRDCLHEFDERLPEIISAMRDKDLLLITADHGNDPTYAGTDHTREY 355
            DFDA YGHRRD  GY   + +FD RLPEII AM++ DLL+ITADHGNDP+Y GTDHTREY
Sbjct: 304 DFDAKYGHRRDVEGYGKAIEDFDGRLPEIIDAMKEDDLLMITADHGNDPSYVGTDHTREY 363

Query: 356 IPLLAYSPSFTGNGLIPVGHFADISATVADNFGVDTAMIGESFLQDLV 403
            IPL+ +S SF    ++PVGHFADISAT+A+NF V  A  GESFL  LV
Sbjct: 364 IPLVIFSKSFKEPKVLPVGHFADISATIAENFSVKKAQTGESFLDALV 411
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2739> which encodes the amino acid sequence <SEQ ID 2740>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0185(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 348/402 (86%), Positives = 374/402 (92%)

Query: 1   MSQFDRIHLVVLDSVGIGAAPDANDFVNAGVPDGASDTLGHISKTVGLAVPNMAKIGLGN 60
            MS+F+RIHLVVLDSVGIGAAPDA+ F NAGV D  SDTLGHIS+  GL+VPNMAKIGLGN
Sbjct: 1   MSKFNRIHLVVLDSVGIGAAPDADKFFNAGVADTDSDTLGHISEAAGLSVPNMAKIGLGN 60

Query: 61  IPRPQALKTVPAEENPSGYATKLQEVSLGKDTMTGHWEIMGLNITEPFDTFWNGFPEDII 120
            I RP  LKTVP E+NP+GY TKL+EVSLGKDTMTGHWEIMGLNITEPFDTFWNGFPE+I+
Sbjct: 61  ISRPIPLKTVPTEDNPTGYVTKLEEVSLGKDTMTGHWEIMGLNITEPFDTFWNGFPEEIL 120

Query: 121 TKIEDFSGRKVIREANKPYSGTAVIDDFGPRQMETGELIIYTSADPVLQIAAHEDIIPLE 180
            TKIE+FSGRK+IREANKPYSGTAVIDDFGPRQMETGELI+YTSADPVLQIAAHEDIIP+E
Sbjct: 121 TKIEEFSGRKIIREANKPYSGTAVIDDFGPRQMETGELIVYTSADPVLQIAAHEDIIPVE 180

Query: 181 ELYRICEYARSITMERPALLGRIIARPYVGEPGNFTRTANRHDYAVSPFEDTVLNKLDQA 240
            ELY+ICEYARSIT+ERPALLGRIIARPYVG+PGNFTRTANRHDYAVSPF+DTVLNKL  A
Sbjct: 181 ELYKICEYARSITLERPALLGRIIARPYVGDPGNFTRTANRHDYAVSPFQDTVLNKLADA 240
```

```
Query: 241 GIDTYAVGKINDIFNGSGINHDMGHNKSNSHGIDTLIKTMGLSEFEKGFSFTNLVDFDAL 300
           G+ TYAVGKINDIFNGSGI +DMGHNKSNSHGIDTLIKT+ L EF KGFSFTNLVDFDA
Sbjct: 241 GVPTYAVGKINDIFNGSGITNDMGHNKSNSHGIDTLIKTLQLPEFTKGFSFTNLVDFDAN 300

Query: 301 YGHRRDPHGYRDCLHEFDERLPEIISAMRDKDLLLITADHGNDPTYAGTDHTREYIPLLA 360
           +GHRRDP GYRDCLHEFD RLPEII+ M++ DLLLITADHGNDPTYAGTDHTREYIPLLA
Sbjct: 301 FGHRRDPEGYRDCLHEFDNRLPEIIANMKEDDLLLITADHGNDPTYAGTDHTREYIPLLA 360

Query: 361 YSPSFTGNGLIPVGHFADISATVADNFGVDTAMIGESFLQDL 402
           YS SFTGNGLIP GHFADISATVA+NFGVDTAMIGESFL  L
Sbjct: 361 YSVSFTGNGLIPQGHFADISATVAENFGVDTAMIGESFLSHL 402
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 900

A DNA sequence (GBSx0955) was identified in *S.agalactiae* <SEQ ID 2741> which encodes the amino acid sequence <SEQ ID 2742>. This protein is predicted to be unnamed protein product (mtaP). Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.44   Transmembrane  215 - 231 ( 215 - 231)

----- Final Results -----
               bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2743> which encodes the amino acid sequence <SEQ ID 2744>. Analysis of this protein sequence reveals the following:

```
     Possible site: 36

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.44   Transmembrane  215 - 231 ( 215 - 231)

----- Final Results -----
               bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 225/269 (83%), Positives = 248/269 (91%)

Query: 1   MTLLEKINETRDFLQAKGVTAPEFGLILGSGLGELAEEIENPIVVDYADIPNWGQSTVVG 60
           M+L+ KINET+DFL  KG+  PEFGLILGSGLGELAEE+EN IV+DYADIPNWG+STVVG
Sbjct: 1   MSLMTKINETKDFLVTKGIETPEFGLILGSGLGELAEEVENAIVIDYADIPNWGKSTVVG 60

Query: 61  HAGKLVYGDLSGRKVLALQGRFHFYEGNTMEVVTFPVRIMRALACHSVLVTNAAGGIGYG 120
           HAGKLVYGDL+GRKVLALQGRFHFYEGN +EVVTFPVR+M+AL C  VLVTNAAGGIGYG
Sbjct: 61  HAGKLVYGDLAGRKVLALQGRFHFYEGNPLEVVTFPVRVMKALGCEGVLVTNAAGGIGYG 120

Query: 121 PGTLMLIKDHINMIGTNPLIGENLEEFGPRFPDMSDAYTATYRQKAHQIAEKQNIKLEEG 180
           PGTLM I DHINM G NPLIGENL+EFGPRFPDMSDAYT  YR KAH++AEK NIKLE+G
Sbjct: 121 PGTLMAITDHINMTGNNPLIGENLDEFGPRFPDMSDAYTKVYRNKAHEVAEKMNIKLEDG 180

Query: 181 VYLGVSGPTYETPAEIRAFQTMGAQAVGMSTVPEVIVAAHSGLKVLGISAITNFAAGFQS 240
           VY+G++GPTYETPAEIRAF+ +GA AVGMSTVPEVIVAAHSGLKVLGISAITNFAAGFQS
Sbjct: 181 VYMGLTGPTYETPAEIRAFKVLGADAVGMSTVPEVIVAAHSGLKVLGISAITNFAAGFQS 240

Query: 241 ELNHEEVVEVTQRIKEDFKGLVKSLVAEL 269
```

```
           ELNHEEVVEVTQ IKEDFKGLVK+++AEL
Sbjct: 241 ELNHEEVVEVTQHIKEDFKGLVKAILAEL 269
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 901

A DNA sequence (GBSx0956) was identified in *S.agalactiae* <SEQ ID 2745> which encodes the amino acid sequence <SEQ ID 2746>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -9.34   Transmembrane  266 - 282 ( 263 - 289)
   INTEGRAL    Likelihood = -8.97   Transmembrane  231 - 247 ( 229 - 253)
   INTEGRAL    Likelihood = -7.70   Transmembrane  356 - 372 ( 352 - 376)
   INTEGRAL    Likelihood = -7.32   Transmembrane  303 - 319 ( 297 - 326)
   INTEGRAL    Likelihood = -5.57   Transmembrane  337 - 353 ( 334 - 355)
   INTEGRAL    Likelihood = -5.57   Transmembrane  391 - 407 ( 387 - 409)
   INTEGRAL    Likelihood = -2.44   Transmembrane  177 - 193 ( 177 - 193)
   INTEGRAL    Likelihood = -1.01   Transmembrane  159 - 175 ( 159 - 175)
   INTEGRAL    Likelihood = -0.43   Transmembrane  198 - 214 ( 196 - 215)

----- Final Results -----
               bacterial membrane --- Certainty=0.4736(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9883> which encodes amino acid sequence <SEQ ID 9884> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAD53928 GB:AF179611 chloride channel protein [Zymomonas
           mobilis]
 Identities = 121/410 (29%), Positives = 213/410 (51%), Gaps = 19/410 (4%)

Query: 14  VKFMIAVLFMTVMAGVGAILMHYVLMFTEWLAFGDSRENTLSLLN------SVTPIKRVL 67
           +++ +A L +  + G+G +L+ ++L   + +A+G S ++ +S  +      + +P++R+
Sbjct: 3   IRYGLACLAVGCLTGLGGMLLSWILHAVQHIAYGYSLQHVISEESFLKGSMAASPLRRLE 62

Query: 68  SLTLVSFLASLSWYYLQIKPKQITSIKQQVVFKDFSVKKSPYWLHIGHAFLQLIYVGTGG 127
            L     +    W  L+     + SI Q V       +  P+W  I H  LQ++ VG G
Sbjct: 63  VLVFCGAVVGGGWGLLRHFGSPLVSITQAVAANK---RVMPFWTTIIHVLLQIVTVGLGS 119

Query: 128 PIGKEGAPREFGAINAGKISDLLALKVLDKRLLIISGAAAGLSAVYQVPLASVFFAFETL 187
           P+G+E APRE G++   + +    L    +R+L+  GA AG ++VY VPL+   FA E L
Sbjct: 120 PLGREVAPRELGSLIGERFAFWGGLSENQRRILVACGAGAGFASVYNVPLSGALFALEAL 179

Query: 188 ALGISLKNIVTLLASTFGAASIAQLVISTAPLYHISKMSLNSQSLAFMFLIVLCVTPI-- 245
            +  +   ++  L ++  +A +A +++  + +YH+    ++++    + L+ L   PI
Sbjct: 180 LMTWASPVVIVALLTSALSARMAWILLGNSMVYHVPAWPVDTR----LMLLALLAGPIFG 235

Query: 246 --AISFRYLNQKVTERRIK-NIKILLSLPVVSLIVSVLSIVYPQILGNGNALVQEVFKGT 302
             A  FR+ +QK+T  RIK N ++ L   +    + +LS+ +P+ILGNG   V   F
Sbjct: 236 IAAHYFRFWSQKITASRIKDNRRLALVAILCFAAIGLLSMWFPEILGNGKGPVSLAFNDN 295

Query: 303 TVSLIA-ILVVLKMIATLSTLYAGAYGGILTPSFSIGACLGFLLASISIPLLPHISIVTS 361
              + A  L   K++A    L+AGAYGG+LTP  S GA L  ++  +    LP + I
Sbjct: 296 LSGMKAGELFCFKILAVFLALWAGAYGGLLTPGISFGALLAVVIGHLWNMWLPPVPIGAF 355

Query: 362 MLVGAAIFLAITMRAPLTAVGLVISFTGQSVITIVPLTIAVLFATAYDYF 411
            ++G A FLA +M+ P+TA+ LVI F       ++P+  AV  + A   F
Sbjct: 356 AIIGGAAFLASSMKMPITAMALVIEFARTGHDFLIPIAFAVAGSIAISQF 405
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2747> which encodes the amino acid sequence <SEQ ID 2748>. Analysis of this protein sequence reveals the following:

```
          Possible site: 13
     >>> Seems to have no N-terminal signal sequence
        INTEGRAL    Likelihood = -5.41   Transmembrane  247 - 263  ( 245 - 267)
        INTEGRAL    Likelihood = -5.15   Transmembrane  326 - 342  ( 323 - 345)
        INTEGRAL    Likelihood = -5.04   Transmembrane  411 - 427  ( 407 - 429)
        INTEGRAL    Likelihood = -4.94   Transmembrane   39 -  55  (  34 -  59)
        INTEGRAL    Likelihood = -4.46   Transmembrane  284 - 300  ( 282 - 307)
        INTEGRAL    Likelihood = -3.45   Transmembrane  380 - 396  ( 376 - 400)
        INTEGRAL    Likelihood = -2.13   Transmembrane  185 - 201  ( 184 - 201)
        INTEGRAL    Likelihood = -2.02   Transmembrane   88 - 104  (  87 - 105)
        INTEGRAL    Likelihood = -1.12   Transmembrane  350 - 366  ( 350 - 367)

     ----- Final Results -----
                   bacterial membrane --- Certainty=0.3166(Affirmative) < succ>
                    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF41386 GB:AE002449 chloride channel protein-related protein
           [Neisseria meningitidis MC58]
 Identities = 137/373 (36%), Positives = 201/373 (53%), Gaps = 23/373 (6%)

Query: 59  IHLIQSLSFGFSQG----SFSTMIASVPPQRRALSLLFAGLLAGLGWHLLAKKGKDIQSI 114
            +H IQ  ++G+       SF   +A     RR   L   G +AG GW LL + GK    I
Sbjct: 1   MHFIQHTAYGYGADGVYTSFREGVAQASGMRRVAVLTLCGAVAGSGWWLLKRFGKPQIEI 60

Query: 115 QQIIQDDISFSPW-TQFWHGWLQLTTVSMGAPVGREGASREVAVTLTSLWSQRCNLSKAD 173
            +  ++  +   P+ T  +H  LQ+ TV +G+P+GRE A RE+         +R  L + +
Sbjct: 61  KAALKQPLQGLPFLTTVFHVLLQIITVGLGSPLGREVAPREMTAAFAFAGGKRLGLDEGE 120

Query: 174 QKLLLACASGAALGAVYNAPLATILFILEAILNRWSLKNIYAACLTSYVAVETVALLQGR 233
             +LL+ACASGA L AVYN PLA+ LFILEA+L  W+ + + AA LTS +A     +  G
Sbjct: 121 MRLLIACASGAGLAAVYNVPLASTLFILEAMLGVWTQQAVAAALLTSVIATAVARI--GL 178

Query: 234 HEIQYLMPQQHWTLGT--LIGSVLAGLILSLFAHAYKHLLKHLPKADAKSQWFIPKVLIA 291
             ++Q   P  + T+ T  L  S + G IL + A  ++   +  P     +   IP  +
Sbjct: 179 GDVQQYHP-ANLTVNTSLLWFSAVIGPILGVAAVFFQRTAQKFPFIKRDNIKIIPLAVCM 237

Query: 292 FSLIAGLSIFFPEILGNGKAG--LLF-FLHEEPH---LSYISWLLVAKAVAISLVFASGA 345
            F+LI  +S++FPEILGNGKAG  L F  L + H   L+ + WL+V  A+A+      GA
Sbjct: 238 FALIGVISVWFPEILGNGKAGNQLTFGGLTDWQHSLGLTAVKWLVVLMALAV------GA 291

Query: 346 KGGKIAPSMMLGGASGLLLAILSQYLIPLSLSNTLAIMVGATIFLGVINKIPLAAPVFLV 405
             GG I PSMMLG       A     + P  +S+  A +VGA +FLGV  K+PL A  F++
Sbjct: 292 YGGLITPSMMLGSTIAFAAATAWNSVFP-EMSSESAAIVGAAVFLGVSLKMPLTAIAFIL 350

Query: 406 EITGQSLLMIIPL 418
            E+T   + +++PL
Sbjct: 351 ELTYAPVALLMPL 363
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 131/415 (31%), Positives = 215/415 (51%), Gaps = 9/415 (2%)

Query: 2   LNFKMVSRLYYAVKFMIAVLFMT-VMAGVGAILMHYVLMFTEWLAFGDSRENTLSLLNSV 60
            LNF   S +      +   LF+T + AG+ A ++   +   + L+FG S+ +  +++ SV
Sbjct: 22  LNFCYNSLMKRHFLLLTFYLFLTGLTAGLVAFILTKAIHLIQSLSFGFSQGSFSTMIASV 81

Query: 61  TPIKRVLSLTLVSFLASLSWYYLQIKPKQITSIKQQVVFKDFSVKKSPYWLHIGHAFLQL 120
             P +R LSL     LA L W+ L  K K I SI QQ++  D S   SP W    H +LQL
Sbjct: 82  PPQRRALSLLFAGLLAGLGWHLLAKKGKDIQSI-QQIIQDDISF--SP-WTQFWHGWLQL 137

Query: 121 IYVGTGGPIGKEGAPREFGAINAGKISDLLALKVLDKRLLIISGAAAGLSAVYQVPLASV 180
```

```
             V  G P+G+EGA RE         S    L   D++LL+   + A L AVY  PLA++
Sbjct: 138 TTVSMGAPVGREGASREVAVTLTSLWSQRCNLSKADQKLLLACASGAALGAVYNAPLATI 197

Query: 181 FFAFETLALGISLKNIVTLLASTFGAASIAQLVISTAPL-YHISKMSLNSQSLAFMFLIV 239
             F  E +    SLKNI     +++ A     L+     + Y + +      +L    L
Sbjct: 198 LFILEAILNRWSLKNIYAACLTSYVAVETVALLQGRHEIQYLMPQQHWTLGTLIGSVLAG 257

Query: 240 LCVTPIAISFRYLNQKVTERRIKNIKILLSLPVVSLIVSVLSIVYPQILGNGNA-LVQEV 298
            L ++  A ++++L + + +   K+   +  + +   +++ LSI +P+ILGNG A L+  +
Sbjct: 258 LILSLFAHAYKHLLKHLPKADAKSQWFIPKVLIAFSLIAGLSIFFPEILGNGKAGLLFFL 317

Query: 299 FKGTTVSLIAILVVLKMIATLSTLYAGAYGGILTPSFSIGACLGFLLASISIPLLP-HIS 357
             +   +S I+ L+V K +A      +GA GG + PS  +G   G LLA +S  L+P  +S
Sbjct: 318 HEEPHLSYISWLLVAKAVAISLVFASGAKGGKIAPSMMLGGASGLLLAILSQYLIPLSLS 377

Query: 358 IVTSMLVGAAIFLAITMRAPLTAVGLVISFTGQSVITIVPLTIA-VLFATAYDYF 411
               +++VGA IFL +  + PL A   ++  TGQS++ I+PL +A ++F  +Y ++
Sbjct: 378 NTLAIMVGATIFLGVINKIPLAAPVFLVEITGQSLLMIIPLALANLIFYFSYQFY 432
```

A related GBS gene <SEQ ID 8683> and protein <SEQ ID 8684> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
SRCFLG: 0
McG: Length  of  UR:   19
     Peak Value of UR:   2.96
     Net Charge of CR: 2
McG: Discrim Score:      9.64
GvH: Signal Score (-7.5): 1.15
     Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 27
ALOM program   count: 9 value:  -9.34 threshold:  0.0
   INTEGRAL    Likelihood = -9.34   Transmembrane  261 - 277 ( 258 - 284)
   INTEGRAL    Likelihood = -8.97   Transmembrane  226 - 242 ( 224 - 248)
   INTEGRAL    Likelihood = -7.70   Transmembrane  351 - 367 ( 347 - 371)
   INTEGRAL    Likelihood = -7.32   Transmembrane  298 - 314 ( 292 - 321)
   INTEGRAL    Likelihood = -5.57   Transmembrane  332 - 348 ( 329 - 350)
   INTEGRAL    Likelihood = -5.57   Transmembrane  386 - 402 ( 382 - 404)
   INTEGRAL    Likelihood = -2.44   Transmembrane  172 - 188 ( 172 - 188)
   INTEGRAL    Likelihood = -1.01   Transmembrane  154 - 170 ( 154 - 170)
   INTEGRAL    Likelihood = -0.43   Transmembrane  193 - 209 ( 191 - 210)
   PERIPHERAL  Likelihood =  1.22      61
 modified ALOM score:   2.37
icml HYPID: 7  CFP: 0.474

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.4736(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00327(340 - 1533 of 1869)
GP|5834362|gb|AAD53928.1|AF179611_12|AF179611(3  -  405  of  425)  chloride  channel  protein
{Zymomonas mobilis}
%Match = 14.7
%Identity = 30.2  %Similarity = 56.1
Matches = 121  Mismatches = 169  Conservative Sub.s = 104

270       300       330       360       390       420       450       468
RSLKLLSVLKKISRD*LNH*LLNFKMVSRLYYAVKFMIAVLFMTVMAGVGAILMHYVLMFTEWLAFGDSRENTLS----L
                                 ::: :| | :  : |:| :|: ::|   : :|:| | :: :|    |
                               MKIRYGLACLAVGCLTGLGGMLLSWILHAVQHIAYGYSLQHVISEESFL
                                       10        20        30        40
```

```
492       522       552       582       612       642       672       702
LNSV--TPIKRVLSLTLVSFLASLSWYYLQIKPKQITSIKQQVVFKDFSVKKSPYWLHIGHAFLQLIYVGTGGPIGKEGA
  |:  :|::|:  | :   :    |  |:     : || | |       :  |:|  | | :||:: || | |:|:| |
KGSMAASPLRRLEVLVFCGAVVGGGWGLLRHFGSPLVSITQAVAANK---RVMPFWTTIIHVLLQIVTVGLGSPLGREVA
         60        70        80        90          100       110       120

732       762       792       822       852       882       912       942
PREFGAINAGKISDLLALKVLDKRLLIISGAAAGLSAVYQVPLASVFFAFETLALGISLKNIVTLLASTFGAASIAQLVI
|||:|::   : :    |    :|:|:  || ||:::|| |||:  :||:| | :  :   ::  | ::  :| :| :::
PRELGSLIGERFAFWGGLSENQRRILVACGAGAGFASVYNVPLSGALFALEALLMTWASPVVIVALLTSALSARMAWILL
           140       150       160       170       180       190       200

972      1002      1032      1059      1089      1119      1146      1176
STAPLYHISKMSLNSQSLAFMFLIVLCVTPIAIS-FRYLNQKVTERRIKNIKILLSLPVVSLI-VSVLSIVYPQILGNGN
  : :||:    :::: | :: |:   :  ||   ||: :||:|  |||: : |  : :: : :  : :||: :|:|||||
GNSMVYHVPAWPVDTR-LMLLALLAGPIFGIAAHYFRFWSQKITASRIKDNRRLALVAILCFAAIGLLSMWFPEILGNGK
           220        230       240       250       260       270       280

1206      1233      1263      1293      1323      1353      1383      1413
ALVQEVFKGTTVSLIA-ILVVLKMIATLSTLYAGAYGGILTPSFSIGACLGFLLASISIPLLPHISIVTSMLVGAAIFLA
  |   |      : |  |  :|::| :  |:||||||:|||  | || |  ::  :    || : |    ::| | |||
GPVSLAFNDNLSGMKAGELFCFKILAVFLALWAGAYGGLLTPGISFGALLAVVIGHLWNMWLPPVPIGAFAIIGGAAFLA
            300       310       320       330       340       350       360

1443      1473      1503      1533      1563      1593      1623      1653
ITMRAPLTAVGLVISFTGQSVITIVPLTIAVLFATAYDYFIRKMRSLYVNPY*SKTR*NCR*NFTSRRSTPCEIYCREFF
 :|: |:||: ||| |       ::|:  ||  : |   |
SSMKMPITAMALVIEFARTGHDFLIPIAFAVAGSIAISQFYDQKKQPKTASKSVISHLGG
            380       390       400       410       420
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 902

A DNA sequence (GBSx0957) was identified in *S.agalactiae* <SEQ ID 2749> which encodes the amino acid sequence <SEQ ID 2750>. This protein is predicted to be purine nucleoside phosphorylase , fragment (deoD-1). Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2384(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAC18350 GB:Y17900 putative purine-nucleotide phosphorylase
          [Streptococcus salivarius]
 Identities = 200/236 (84%), Positives = 219/236 (92%)

Query: 1   MSIHIEAKQGEIADKILLPGDPLRAKFIAENFLEDAVCFNTVRNMFGYTGTYKGHRVSVM 60
           MSIHI AKQGEIADKILLPGDPLRAKFIAENFLEDAVCFN VRNMFGYTGTYKG RVSVM
Sbjct: 1   MSIHIAAKQGEIADKILLPGDPLRAKFIAENFLEDAVCFNEVRNMFGYTGTYKGERVSVM 60

Query: 61  GTGMGMPSISIYARELIVDYGVKTLIRVGTAGAINPDIHVRELVLAQAAATNSNIIRNDW 120
           GTGMGMPSISIYARELIVDYGVK LIRVGTAG++N D+HVRELVLAQAAATNSNIIRNDW
Sbjct: 61  GTGMGMPSISIYARELIVDYGVKKLIRVGTAGSLNEDVHVRELVLAQAAATNSNIIRNDW 120

Query: 121 PEFDFPQIADFKLLDKAYHIAKEMDITTHVGSVLSSDVFYSNQPDRNMALGKLGVHAIEM 180
           P++DFPQIA+F LLDKAYHIAK   +TTHVG+VLSSDVFYSN  ++N+ LGK GV A+EM
Sbjct: 121 PQYDFPQIANFNLLDKAYHIAKNFGMTTHVGNVLSSDVFYSNYFEKNIELGKWGVKAVEM 180
```